2019-03-01

Use of Species Distribution Modeling in the Deep Sea

Kenchington, E

Use of Species Distribution Modeling in the Deep Sea

E. Kenchington, O. Callery, F. Davidson, A. Grehan, T. Morato, J. Appiott, A. Davis, P. Dunstan, C. Du Preez, J. Finney, J.M. González-Irusta, K. Howell, A. Knudby, M. Lacharité, J. Lee, F. J. Murillo, L. Beazley, J.M. Roberts, M. Roberts, C. Rooper, A. Rowden, E. Rubidge, R. Stanley, D. Stirling, K.R. Tanaka, J. Vanhatalo, B. Weigel, S. Woolley and C. Yesson

Ocean and Ecosystem Sciences Division
Maritimes Region
Fisheries and Oceans Canada

Bedford Institute of Oceanography
PO Box 1006
Dartmouth, Nova Scotia
Canada B2Y 4A2

2019

Canadian Technical Report of Fisheries and Aquatic Sciences 3296

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of
Fisheries and Aquatic Sciences 3296

2019

USE OF SPECIES DISTRIBUTION MODELING IN THE DEEP SEA

by

E. Kenchington[1], O. Callery[2], F. Davidson[3], A. Grehan[2], T. Morato[4], J. Appiott[5], A. Davis[6], P. Dunstan[7], C. Du Preez[8], J. Finney[9], J.M. González-Irusta[4], K. Howell[10], A. Knudby[3], M. Lacharité[11], J. Lee[5], F.J. Murillo[1], L. Beazley[1], J.M. Roberts[12], M. Roberts[6], C. Rooper[13], A. Rowden[14], E. Rubidge[8], R. Stanley[1], D. Stirling[15], K.R. Tanaka[16], J. Vanhatalo[17], B. Weigel[17], S. Woolley[7] and C. Yesson[18]

[1] Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS, Canada
[2] National University of Ireland, Galway, Ireland
[3] University of Ottawa, Ottawa, Ontario, Canada
[4] Institute of Marine Research, Azores, Portugal
[5] Secretariat of the Convention on Biological Diversity, Montreal, Quebec, Canada
[6] University of Bangor, Bangor, Wales, United Kingdom
[7] CSIRO Oceans and Atmosphere, Hobart, Tasmania, Australia
[8] Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, British Columbia, Canada
[9] Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, British Columbia, Canada
[10] University of Plymouth, Plymouth, England, United Kingdom
[11] Nova Scotia Community College, Dartmouth, Nova Scotia, Canada
[12] University of Edinburgh, Edinburgh, Scotland, United Kingdom
[13] Alaska Fisheries Science Center, Seattle, Washington, United States of America
[14] National Institute of Water and Atmospheric Research, Wellington, New Zealand
[15] Marine Scotland Science, Aberdeen, Scotland, United Kingdom
[16] Princeton University, Princeton, New Jersey, United States of America
[17] University of Helsinki, Helsinki, Finland
[18] Zoological Society of London, London, England, United Kingdom

Cat. N ° Fs97-6/3296E-PDF ISBN 978-0-660-29721-7 ISSN 1488-5379

Correct citation for this publication:

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C., Finney, J., González-Irusta, J.M., Howell, K., Knudby, A., Lacharité, M., Lee, J., Murillo, F.J., Beazley, L., Roberts, J.M., Roberts, M., Rooper, C., Rowden, A., Rubidge, E., Stanley, R., Stirling, D., Tanaka, K.R., Vanhatalo, J., Weigel, B., Woolley, S. and Yesson, C. 2019. Use of Species Distribution Modeling in the Deep Sea. Can. Tech. Rep. Fish. Aquat. Sci. 3296: ix + 76 p.

TABLE OF CONTENTS

ABSTRACT

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C., Finney, J., González-Irusta, J.M., Howell, K., Knudby, A., Lacharité, M., Lee, J., Murillo, F.J., Beazley, L., Roberts, J.M., Roberts, M., Rooper, C., Rowden, A., Rubidge, E., Stanley, R., Stirling, D., Tanaka, K.R., Vanhatalo, J., Weigel, B., Woolley, S. and Yesson, C. 2019. Use of Species Distribution Modeling in the Deep Sea. Can. Tech. Rep. Fish. Aquat. Sci. 3296: ix + 76 p.

In the last two decades the use of species distribution modeling (SDM) for the study and management of marine species has increased dramatically. The availability of predictor variables on a global scale and the ease of use of SDM techniques have resulted in a proliferation of research on the topic of species distribution in the deep sea. Translation of research projects into management tools that can be used to make decisions in the face of changing climate and increasing exploitation of deep-sea resources has been less rapid but necessary. The goal of this workshop was to discuss methods and variables for modeling species distributions in deep-sea habitats and produce standards that can be used to judge SDMs that may be useful to meet management and conservation goals. During the workshop, approaches to modeling and environmental data were discussed and guidelines developed including the desire that 1) environmental variables should be chosen for ecological significance *a priori*; 2) the scale and accuracy of environmental data should be considered in choosing a modeling method; 3) when possible proxy variables such as depth should be avoided if causal variables are available; 4) models with statistically robust and rigorous outputs are preferred, but not always possible; and 5) model validation is important. Although general guidelines for SDMs were developed, in most cases management issues and objectives should be considered when designing a modeling project. In particular, the trade-off between model complexity and researcher's ability to communicate input data, modeling method, results and uncertainty is an important consideration for the target audience.

RÉSUMÉ

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C., Finney, J., González-Irusta, J.M., Howell, K., Knudby, A., Lacharité, M., Lee, J., Murillo, F.J., Beazley, L., Roberts, J.M., Roberts, M., Rooper, C., Rowden, A., Rubidge, E., Stanley, R., Stirling, D., Tanaka, K.R., Vanhatalo, J., Weigel, B., Woolley, S. and Yesson, C. 2019. Recours à la modélisation de la répartition des espèces en haute mer. Can. Tech. Rep. Fish. Aquat. Sci. 3296: ix + 76 p.

Au cours des deux dernières décennies, le recours à la modélisation de la répartition des espèces pour étudier et gérer les espèces marines a considérablement augmenté. La disponibilité des variables prédictives à l'échelle mondiale et la convivialité de ces techniques de modélisation ont entraîné la multiplication des recherches sur la répartition des espèces en haute mer. La traduction des projets de recherche en outils de gestion pouvant servir à prendre des décisions dans le contexte des changements climatiques et de l'exploitation accrue des ressources en haute mer est moins rapide, quoique nécessaire. Cet atelier visait à discuter des méthodes et variables pour la modélisation de la répartition des espèces dans les habitats en haute mer, et à établir des normes pour évaluer les méthodes de modélisation pouvant aider à atteindre les objectifs en matière de gestion et de conservation. Pendant l'atelier, les approches envers la modélisation et les données environnementales ont fait l'objet de discussions, et des lignes directrices ont été élaborées. Celles-ci comprenaient les caractéristiques souhaitées qui suivent : 1) les variables environnementales devraient être choisies selon leur importance écologique a priori; 2) l'ampleur et l'exactitude des données environnementales devraient être prises en compte durant la sélection d'une méthode de modélisation; 3) dans la mesure du possible, les variables substitutives, comme la profondeur, doivent être évitées si des variables causales sont disponibles; 4) les modèles dont les résultats sont statistiquement solides et rigoureux sont privilégiés, mais leur utilisation n'est pas toujours possible; 5) la validation du modèle est importante. Même si des lignes générales sur la modélisation de la répartition des espèces ont été mises au point, les objectifs et enjeux de gestion devraient généralement être pris en compte pendant la conception d'un projet de modélisation. En particulier, le compromis entre la complexité du modèle et la capacité du chercheur à communiquer les données d'entrée, la méthode de modélisation, les résultats et les incertitudes sont des facteurs importants pour le public cible.

1. INTRODUCTION

The use of species distribution modeling (SDM; also known as habitat suitability modeling or HSM) in marine ecology has grown enormously in recent years with the accumulation of appropriate response and predictor data sets, and software that is freely available and often global in scope. Exploitation of the deep sea is increasing and diversifying, while knowledge of the actual distribution of deep-sea biodiversity is still poor due to the limited spatial footprint of sampling to date. There is now a great need for robust species distribution modeling that can inform decision-making and anticipate the influence of global change. In particular, to support the *Voluntary Specific Workplan on Biodiversity in Cold-water Areas within the Jurisdictional Scope of the Convention* as adopted at CBD COP13. However, the variables that may be useful predictors for species and communities living in shallow waters and on the continental shelves may not be appropriate for modeling distributions below 100 m. For example, in the northwest Atlantic temperature and salinity are relatively constant below this depth but are key determinants of distributions at shallower depths. The goal of this workshop was to discuss appropriate variables and methods for modeling of species and communities of deep-sea habitats, including seamounts, and to produce a set of standards for publication that will lead to more comparable work in this field.

This document summarizes the discussions of a workshop that brought together global experts (including from the EU Horizon 2020 research projects ATLAS and SponGES) on species distribution modeling and deep-sea biology to discuss this topic. The main objectives of the workshop were to: 1) showcase existing attempts to develop species distribution modeling in the deep sea; 2) to discuss biological and environmental datasets relevant to the deep sea; 3) discuss temporal and spatial scales relevant for developing species distribution modeling in the deep sea; 4) discuss appropriate modeling tools in the context of data-limited situations, and in the context of single species and joint species modeling; and 5) establish more effective collaborations within the deep-sea modeling community.

Participants were asked to prepare short working papers addressing agenda items and to make ten minute presentations to stimulate discussion. Those presentations are presented in the Extended Abstract section below.

Co-chairs: Ellen Kenchington (DFO, Canada), Anthony Grehan (NUI, Galway, Ireland), Telmo Morato (IMAR, Azores, Portugal)

1.1. Setting the Context: Voluntary Specific Workplan on Biodiversity in Cold-water Areas within the Jurisdictional Scope of the Convention on Biological Diversity

J. Murray Roberts

ATLAS project coordinator, University of Edinburgh, School of GeoSciences, Grant Institute, James Hutton Road, Edinburgh, UK

In 2012 the 11th Conference of the Parties (COP) to the Convention on Biological Diversity (CBD) requested that the CBD Executive Secretary prepare a draft specific workplan on biodiversity and ocean acidification in cold-water areas of the ocean. Developed through a collaborative process involving other governments and relevant organisations, this workplan would build upon elements of a previous workplan on physical degradation and destruction of coral reefs, including cold-water corals (CBD, 2004). The 'cold-water areas' workplan should link closely with relevant work under the CBD, including the description of areas meeting the scientific criteria for Ecologically or Biologically Significant marine Areas (EBSAs) and the UN Food & Agriculture Organisation's (FAO) identification of Vulnerable Marine Ecosystems (VMEs).

This process began when the CBD Executive Secretary issued a notification in May 2015 requesting scientific and technical information and suggestions from Parties, other Governments and relevant organisations. Information was received from Argentina, Australia, Brazil, Canada, Colombia, France, Mexico, New Zealand, the United Kingdom of Great Britain and Northern Ireland (UK), the European Union, the International Atomic Energy Agency, the OSPAR Commission and the UN Division on Ocean Affairs and the Law of the Sea. Alongside updates to the 2014 CBD Technical Series report on ocean acidification (CBD, 2014), these contributions were used to prepare a background document for peer review. Following reviews from Canada, the UK and FAO this background document was considered at the 20th meeting of the CBD's Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA) in April 2016 (CBD, 2016a).

The background document and SBSTTA meeting in April informed discussion at the 13th COP in December 2016 where Decision XIII/11, a 'voluntary specific workplan on biodiversity in cold-water areas within the jurisdictional scope of the Convention', was formally adopted (CBD, 2016b). COP Decision XIII/11 helps set the wider context for the expert workshop on deep-sea SDM organised through the European Union's ATLAS and SponGES projects in May 2018 at the CBD Secretariat in Montreal, Canada. Compared to other biomes, the need for reliable species distribution modeling in the deep sea is particularly pressing. This need is because the vast extent, remoteness, technical challenge and expense of deep-sea research greatly limit the amount of species and habitat distribution data available. Reliable deep-sea species distribution models offer a potentially valuable resource to managers struggling to develop policies to manage rapidly increasing human activities in the deep ocean, including on-going fisheries and hydrocarbon extraction through to possible future deep-sea mining. However, while the potential of species distribution modeling is great, there are many issues to be worked through to optimise underlying species occurrence/absence information, environmental data and the modeling approaches applied – issues addressed at the Montreal workshop and described in this report.

The background document underpinning COP Decision XIII/11 considered the biodiversity of deep, cold areas of the global ocean excluding polar regions. Implications of human activities and global change were summarised, and activities to monitor key parameters and relevant policy

instruments were reviewed in that document. It also included analysis of knowledge gaps in both the scientific evidence-base and the policy instruments available. The background document highlighted several key points, many of which are now subject to increased research and management action in several parts of the world:

(a) Cold-water areas sustain ecologically important habitats, such as cold-water corals and sponge fields, which play important functional biological and ecological roles, including supporting rich communities of fish as well as suspension-feeding organisms such as bryozoans and hydroids;
(b) Ocean acidification, increases in ocean temperature and deoxygenation can have significant impacts on biodiversity and ecosystems in cold-water areas, including decreased ocean mixing, changes in nutrient cycling and oxygen supply, community shifts, and impacts on habitat structure and range, and organism physiology;
(c) Ocean acidification, in particular, presents a significant threat to ecosystems in cold-water areas, including through weakening and dissolution of cold-water coral skeletons, and impacts on diverse taxa such as sponges, squid species, pteropods, krill and fish. Because the aragonite saturation horizon is projected to become much shallower by 2100, more cold-water ecosystems and habitats are expected to be exposed to the impacts of ocean acidification in the coming years;
(d) There are existing and potential pressures on biodiversity in cold-water areas from anthropogenic sources, including destructive fishing practices, deep-sea marine mining, hydrocarbon exploitation, shipping and bioprospecting, as well as impacts related to the accumulation of plastic microfibers and other pollutants;
(e) Although knowledge on biodiversity and acidification in cold-water areas is growing and global monitoring of ocean acidification is increasing, there is a need for further research in this area, including on the interactions among species within food webs, impacts of ocean acidification on different life stages of cold-water organisms, impacts of multiple stressors on biodiversity and ecosystems, the goods and services they provide and variability in the response by organisms to various pressures;
(f) There is a need to address pressures through targeted policy and management responses at the global, regional and national levels, supported by identification of specific areas that are of high biological and ecological importance.

Decision XIII/11 encourages Parties to the CBD, other Governments and competent intergovernmental organizations to:

(a) Avoid, minimize and mitigate the impacts of global and local stressors, and especially the combined and cumulative effects of multiple stressors;

(b) Maintain and enhance the resilience of ecosystems in cold-water areas in order to contribute to the achievement of Aichi Biodiversity Targets 10, 11 and 15, and thereby enable the continued provisioning of goods and services;
(c) Identify and protect refugia sites and areas capable of acting as refugia sites, and adopt, as appropriate, other area-based conservation measures, in order to enhance the adaptive capacity of cold-water ecosystems;
(d) Enhance understanding of ecosystems in cold-water areas, **including by improving the ability to predict the occurrence of species and habitats** and to understand their vulnerability to different types of stressors as well as to the combined and cumulative effects of multiple stressors;
(e) Enhance international and regional cooperation in support of national implementation, building on existing international and regional initiatives and creating synergies with various relevant areas of work within the Convention.

The activities proposed in Decision XIII/11 included enhanced policy integration, strengthened management, development and application of Marine Protected Areas, improved monitoring, improved research coordination and capacity building and initiatives that would secure and sustain funding.

The explicit inclusion of predictive mapping approaches was expanded in Annex 3 'Monitoring and research needs' under Subsection 5 to develop or expand upon predictive model research to determine how projected climate change will impact cold-water biodiversity over different time scales:

5.1 Improve ocean carbonate models to understand the temporal and three-dimensional spatial changes in carbonate saturation state and its main drivers, including changing atmospheric CO_2 conditions and ocean currents;
5.2 Document existing gaps in knowledge on global, regional and national scales that limit the predictive power of models;
5.3 Couple ocean carbonate chemistry mapping and oceanographic models to biophysical and ecological information to predict the temporal and spatial variability of acidification impacts in order to help identify areas under the greatest threat as well as possible refugia;
5.4 Optimise habitat modeling to predict key habitats and biodiversity occurrence from seawater carbonate chemistry, oceanographic and water mass modeling and larval dispersal.

The CBD's voluntary specific workplan on biodiversity in cold-water areas is a significant demonstration of how important it has become to not only understand the implications of global change on marine biodiversity and ecosystem function, but to enhance our abilities to monitor and make predictions of how marine ecosystems could change. Given the great complexity and

interconnected nature of deep and open ocean ecosystems the challenge is great but, as the discussions in this workshop demonstrate, the underlying data and techniques required for robust deep-sea predictive species distribution modeling have developed rapidly in the last decade. The onus is now on the marine scientific and policy communities to take up these challenges and deliver integrated assessments that allow us to develop a truly predictive ability to forecast not only where key species occur but how their distributions will change in the future.

Acknowledgements

This paper is a contribution to the European Union's Horizon 2020 research and innovation programme under grant agreement No 678760 (ATLAS). It reflects solely the author's view, and the European Union and CBD cannot be held responsible for any use that may be made of the information contained herein.

References

CBD. 2004. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity. VII/5 Marine and coastal biological diversity. https://www.cbd.int/doc/decisions/cop-07/cop-07-dec-05-en.pdf

CBD. 2014. An updated synthesis of the impacts of ocean acidification on marine biodiversity (Eds: S. Hennige, J.M.Roberts and P. Williamson). Montreal, Technical Series No. 75, 99 pages. https://www.cbd.int/doc/publications/cbd-ts-75-en.pdf

CBD. 2016a. Background document on biodiversity and acidification in cold-water areas. CBD-SBSTTA 20. https://www.cbd.int/doc/meetings/sbstta/sbstta-20/information/sbstta-20-inf-25-en.pdf

CBD. 2016b. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity. Voluntary specific workplan on biodiversity in cold-water areas within the jurisdictional scope of the Convention. CBD-COP 13. https://www.cbd.int/doc/decisions/cop-13/cop-13-dec-11-en.pdf

2. THEME 1 SHOWCASE OF APPROACHES TO DEVELOP SPECIES DISTRIBUTION MODELS IN THE DEEP SEA

The workshop opened with two invited speakers who gave keynote presentations relevant to the workshop theme. The discussion following those presentations is summarized herein.

2.1. Model-Based Thinking for Community Ecology

Piers Dunstan

Commonwealth Scientific and Industrial Research Organisation (CSIRO), Oceans and Atmosphere, Hobart, Tasmania, Australia

One of the key challenges of ecology is to understand the distribution and abundance of species through space and time. Ecologists have been searching for the most parsimonious explanations of these patterns for decades. One of the earliest analytical approaches was developed by Bray and Curtis (1957), and application of similar methods has been extensive in ecology since then. However, recent developments have demonstrated that these distance-based methods may have serious flaws (Warton et al., 2011; Woolley et al., 2017) and alternatives are needed.

Fundamentally, when we are observing ecological systems we are observing individuals from different species – this is the data that can be used in analysis. From these data, we infer assemblages, communities, bioregions and other higher-level properties. Warton et al. (2015a, b) suggests that an appropriate approach to this problem is through model-based ecology. The key concept is that a statistical model can be developed that describes the relationship between the observations (individuals of species) and the higher-level properties of interest (e.g., assemblages, bioregions). These models use latent variables to capture the higher-level properties and use well-described covariates (e.g., environmental) to estimate the functional responses of the latent variables.

Two examples of this approach are Species Archetype Models (SAM, Dunstan et al., 2011; Dunstan et al., 2013) and Regions of Common Profiles (RCP, Foster et al., 2014a; Foster et al., 2017; Hill et al., 2017). Species Archetype Models are multivariate mixture regression models where species are grouped together based on a common response to environmental covariates. Species that have a statistically identical response to environmental gradients are grouped together and a regression model is estimated for each species group. The species groups can then be used to predict the distribution of the group into space using environmental covariates. The model retains information on the species in each group and propagates uncertainty from data to prediction. The models can be estimated with presence only data (Inhomogeneous Poisson Point Models), presence/absence data (binomial models), abundance data (negative binomial models) and biomass data (Tweedie models). Examples of SAMs have been on distribution of species in Western Australia (Woolley et al., 2013) and the response of species to trawl fisheries (Foster et al., 2014b).

RCP models are similar to SAM models, except that the grouping is on sites rather than species. The RCP models group sites based on biological content and see how these groups vary with the environment – these models suppose that assemblages exist. The assumption is that an assemblage can be characterised by its mean expectation of all species (its profile) and that different assemblages have different profiles. This assumption means that different assemblages will have

different expected biological content. Each site contains one assemblage, which we don't observe, but can be described as a mixture of sites (characterised by biological profiles). We label this model as `region of common profiles', but it is also known as a `mixture of experts' model. In a similar way to SAM models, RCP models can handle a range of different data types (presence only data (Inhomogeneous Poisson Point Models), presence/absence data (binomial models), abundance data (negative binomial models) and biomass data (Tweedie models) and propagate uncertainty from the observations to the predictions. Examples of RCPs are the distribution of demersal fish on Kerguelen Plateau (Hill et al., 2017) and predictions where sampling artefacts are present (Foster et al., 2017).

SAM and RCP models often give slightly different predictions as a result of differences in the assumptions. Assemblages often share species with different expected abundances. This property means that there will be multiple species groups from SAMs at any site. However, sites are often the unit of spatial management and distinguishing between sites is an important tool for planning and management decisions. We expect that SAMs and RCPs will give similar results when there are very strong environmental gradients acting across the area of interest, which will cause groups of species to be tightly correlated with each other under specific environmental conditions. As gradients weaken, more species groups should be present at any site.

Future applications of this work will start to link the outputs from both SAMs and RCP with other models types, such as qualitative and quantitative ecosystem models.

References

Bray, J.R., and Curtis, J.T. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol. Monogr. 27(4): 326-349.

Dunstan, P.K., Foster, S.D., and Darnell, R. 2011. Model Based Grouping of Species across Environmental Gradients. Ecol. Modell. 222: 955-963.

Dunstan, P.K., Foster, S.D., Hui, F.K.C., and Warton, D.I. 2013. Finite Mixture of Regression Modeling for high-dimensional count and biomass data in Ecology. J. Agric. Biol. Ecol. Stat. 18(3): 357-375.

Foster, S.D., Givens, G.H., Dornan, G.J., Dunstan, P.K., and Darnell, R. 2014a. Modeling biological regions from multi-species and environmental data. Environmetrics DOI: 10.1002/env.2245

Foster, S.D., Dunstan, P.K., Althaus, F., and Williams, A. 2014b. The cumulative effect of trawl fishing on a multispecies fish assemblage in south-eastern Australia. J. Appl. Ecol. 52:129-139.

Foster, S.D., Hill, N.A., and Lyons, M. 2017. Ecological grouping of survey sites when sampling artefacts are present. Appl. Stat. 66: 1031-1047.

Hill, N.A., Foster, S.D., Duhamel, G., Welsford, D., Koubbi, P. and Johnson, C.R. 2017. Model-based mapping of assemblages for ecology and conservation management: A case study of demersal fish on the Kerguelen Plateau. Divers. Distrib. 23: 1216-1230.

Warton, D.I., Wright, S.T., Wang, Y. 2011. Distance-based multivariate analyses confound location and dispersion effects. Methods Ecol. Evol. 3: 89-101. doi: 10.1111/j.2041-210X.2011.00127.x

Warton, D.I., Foster, S.D., De'ath, G., Stoklosa, J., Dunstan, P.K. 2015a. Model-based thinking for community ecology. Plant Ecol. 216:669–682.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., Hui, F.K.C. 2015b. So Many Variables: Joint Modeling in Community Ecology. Trends Ecol. Evol. 30: 766-799. http://dx.doi.org/10.1016/j.tree.2015.09.007

Woolley, S.N.C., McCallum, A.W., Wilson, R., O'Hara, T.D., Dunstan, P.K. 2013. Fathom out: biogeographical subdivision across the Western Australian continental margin – a multispecies modeling approach. Divers. Distrib. 19: 1506-1517. DOI: 10.1111/ddi.12119

Woolley, S.N.C., Foster, S.D., O'Hara, T.D., Wintle, B.A., Dunstan, P.K. 2017. Characterising uncertainty in generalised dissimilarity models. Methods Ecol. Evol. 8: 985-995. doi: 10.1111/2041-210X.12710

2.1.1. Group Discussion

Following Dr. Dunstan's presentation, there was a discussion comparing and contrasting SAMs and RCP as approaches to the statistical analysis and prediction of multispecies data. It was agreed that each have potential benefits, and the more suitable approach depends on the intended use of the outputs and questions to which answers are being sought. For example, for the purposes of marine spatial planning where we might want to group sites which share common sets of species with common probabilities of occurrence, the RCP approach may be more suitable, while for the purposes of understanding how species and groups of species respond to their environment and various pressures, SAMs may provide more useful predictions. One can expect these two methods (grouping by sites and grouping by species) occasionally will give similar results, for example, when there are strong environmental gradients or strong associations between species. It was agreed that there were obvious benefits to multispecies modeling approaches; not least of which is the reduction in the number of models required to describe an ecosystem comprising large numbers of species. Also, many workshop participants noted that they had often observed multi-species modeling approaches to have better predictive power to describe the distributions of individual species than single species models targeted at that species alone. For example, there are occasions where rare species, for which occurrence data is extremely limited, may be fit to an archetype thereby improving both the overall archetype model as well as its ability to make

predictions for that rare species. It was conceded, however, that where data for a species was so scant as to impede its being fit to an archetype, it may be better to exclude those data entirely.

SAMs offer improvements over previous multi-species modeling as the processes of species assembly and model prediction are performed simultaneously. This approach is in contrast to previously implemented modeling approaches where either 1) species are assembled first and predictions made based on grouped species, or 2) where model predictions are made first and multiple predictions assembled thereafter. The primary advantage of this simultaneous prediction/assembly approach is that users can obtain an estimate of uncertainty – an element that is too often ignored by other modeling approaches despite its importance to the use of model predictions in subsequent decision-making processes.

To facilitate implementation of both SAMs and RCP modeling, two R packages - "SpeciesMix" and "RCPmod" – have been made available on the CRAN repository – these are currently being combined together into a single interface, "EcoMix", to make them more user-friendly and to provide more flexibility. It is envisaged that future work in the area of multispecies modeling will focus on using SAMs combined with mechanistic models with the goal of building whole ecosystem-level models which have both descriptive/predictive and explanatory utility.

2.2. Point Process Framework for Species Distribution Modeling and Joint Species Distribution Models

Jarno Vanhatalo

Department of Mathematics and Statistics and Organismal and Evolutionary Biology Research Programme, University of Helsinki, Finland

Point process framework for species distribution modeling

Herein, I briefly review the point process modeling approach for species distribution modeling and illustrate it with few case examples. The benefits from this approach for deep-sea species distribution modeling could be that they i) cover both occurrence and abundance modeling, ii) allow integration of heterogeneous data from different survey setups (including presence only), and iii) can be straightforwardly extended to current joint species distribution models (JSDM). I will consider log-Gaussian Cox processes (LGCP; Banerjee et al., 2015) whose connection with a common species distribution modeling method, MaxEnt, was shown by Renner and Warton (2013). I will assume spatial LGCP, but the same framework can be extended to spatio-temporal cases.

Let's denote by s as spatial locations (e.g., longitude-latitude coordinates – and possibly depth), and x_s to be the vector of environmental covariates associated with that location (e.g., depth,

temperature etc.). A key component of point process models is the intensity function $\lambda(s, x_s)$ which, loosely speaking, is the probability that one individual is present at location s. The number of animals (abundance) in any fixed area/volume of water, D, is Poisson distributed with expected value $\int_{s \in D} \lambda(s, x_s) \mathrm{d}s$ which is the "total intensity" over D. For example, for benthic animals the intensity function can be directly interpreted as the expected number of animals per unit area (m^2), whereas for pelagic animals this function is interpreted as the expected number of animals per volume of water (m^3). Let's assume that the sampling effort and, hence, observation probability, varies in space, and denote this by $z(s, x_s)$. In this case the expected number of observed individual animals in an area is $\int_{s \in D} \lambda(s, x_s) z(s, x_s) \mathrm{d}s$. Now, the sampling effort can be treated as known or as an uncertain parameter of the model. In many cases the observation probability is not known but can be assumed to be constant, in which case λ corresponds to relative intensity.

In practice LGCPs are typically discretized so that observations are considered to be made over finite areas (e.g., discrete lattice grid), over which the intensity is treated to be constant (Renner and Warton, 2013; Banerjee et al., 2015). Hence, given a set of species observations $y_1, \ldots, y_n$ over a set of n finite areas indexed with locations $\boldsymbol{S} = (s_1, \ldots, s_n)$, the likelihood for an intensity and effort function is

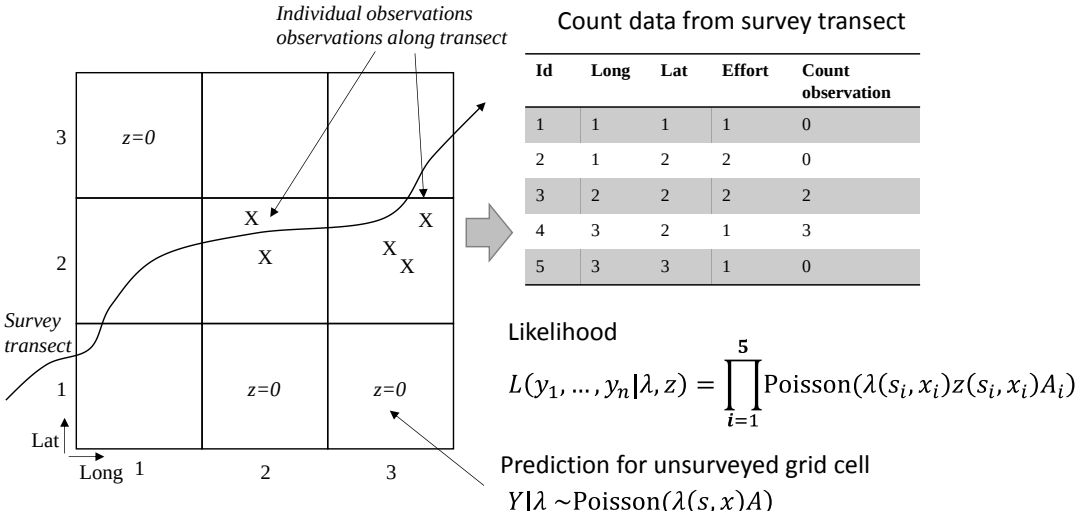
$$L(y_1, \ldots, y_n | \lambda, z) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \mathrm{Poisson}(\lambda(s_i, x_i) z(s_i, x_i) A_i)$$

where s_i is the (center) location of i'th sampling area, and A_i is the size of that area (or volume of water). See Figure 2.2.1 for an illustration. To finalize the species distribution model we build a model for the intensity function which is typically constructed using a log linear model with spatial random effects so that

$$\log \lambda(s, x_s) = f(s, x_s) = \alpha + {x_s}^T \beta + \varphi(s)$$

where the intercept α corresponds to the average log intensity, ${x_s}^T \beta$ describes the effect of environmental covariates and $\varphi(s)$ is a spatial random effect that explains spatially correlated patterns not explained by environmental covariates. The spatial random effect is modelled with zero mean Gaussian process (Banerjee et al., 2015), and it has been shown to improve the explanatory and predictive performance of species distribution models in many studies (Latimer et al., 2006; Vanhatalo et al., 2012; Clark et al., 2014; Kallasvuo et al., 2017). The linear environmental predictors can also be replaced by non-linear functions (e.g., Kallasvuo et al., 2017). If we assume independent random noise per sampling location/area, the Poisson likelihood extends naturally to a Negative-Binomial model that allows for over-dispersion (Vanhatalo et al., 2017).

Hence, from an analysis point of view, simple point process models are generalized linear models. However, the added value is that their interpretation is ecologically meaningful, and allows integration of alternative data sets and prediction of relative biomasses.



Id	Long	Lat	Effort	Count observation
1	1	1	1	0
2	1	2	2	0
3	2	2	2	2
4	3	2	1	3
5	3	3	1	0

Figure 2.2.1. A schematic illustration of a point process model implemented over a lattice grid. The survey data has resulted in a table of count observations at surveyed cells ($z > 0$). After solving the posterior distribution of model parameters, we can predict the species intensity in cells not surveyed.

Bayesian predictive inference and example applications

Relative effect on intensity

Given training data, we can calculate the posterior distribution (Banerjee et al., 2015) of the model parameters, examine the effect of environmental covariates and use the model for making predictions in locations not covered by data. Due to the log-link function, the additive model components have particularly simple interpretation as log relative change in intensity (Vanhatalo et al., 2017). For example, the relative effect to intensity by the spatial random effect is:

$$RE(s) = \frac{\lambda(s,x) - \lambda(s,x;\varphi(s)=0)}{\lambda(s,x;\varphi(s)=0)} = \frac{e^{\alpha + {x_s}^T\beta + \varphi(s)}}{e^{\alpha + {x_s}^T\beta}} - 1 = e^{\varphi(s)} - 1$$

which tells how much larger (>0) or smaller (<0) the density of a species is at location s compared to a density estimate based only on (environmental) covariates. Similarly, the relative effect of the d'th covariate is $RE(x_d) = e^{x_d\beta_d} - 1$ regardless of the other covariates or spatial locations. Vanhatalo et al. (2017) used LGCP to analyze spatial-temporal fluctuations of Crown of Thorns Starfish (COTS) in the Great Barrier Reef, Australia. The log intensity included spatial and spatio-temporal random effects. The discrete sampling areas A_i varied by observations and corresponded to the surveyed reef area. One of the key results was the visualization of the spatial and spatial-

temporal relative effects, which revealed high/low intensity COTS hot/cold spots and COTS outbreak dynamics (Figure 2.2.2).

Bayesian predictive inference and biomass estimates

We can use point process models for (relative) biomass predictions. Kallasvuo et al. (2017) analyzed larval areas of four commercially important coastal fish species. Each of their observations corresponded to number of fish larvae in a volume of water, V_i, sampled by sampling nets and, hence, the intensity function $\lambda(s,t,x)$ corresponded to relative larvae density (number per m^3 multiplied by catch probability). Based on the relative density predictions, Kallasvuo et al. (2017) divided the study region into three classes: *important regions* (the highest density areas producing 80% of total number of larvae), *suitable regions* (regions with >50% probability to observe larvae) and *non-suitable regions* (regions with <50% probability to observe larvae). In practice this analysis was done by predicting λ in approximately 15 million 50 m × 50 m grid cells and calculating the total (relative) number of larvae in the study region, $N_{\text{tot}} = \sum V\lambda(s_i, x_{s_i})$ where the sum goes through all the grid cells and V is the volume of water (surface layer) in the grid cell. The highest intensity grid cells, whose total relative number of larvae was $0.8 \times N_{\text{tot}}$, comprise the important larval production region. One of the key findings of their study was that the important regions can be much smaller in size than regions suitable for larval production. Moreover, the difference between important and suitable areas varied dramatically between species, which shows that using only presence/absence models to predict species distributions might give a biased view of the important areas (see Figure 2.2.3).

Accounting for unequal sampling effort

For a last example, we consider integration of heterogeneous data and modeling the survey effort. Mäkinen and Vanhatalo (2018) conducted marine mammal distribution modeling in the Arctic using heterogeneous data collected from earlier publications and free data sources. The data were comprised of species sightings made during survey cruises, but the exact survey protocol during the surveys was unknown, for which reason they also explicitly modeled the survey effort, z. In their model, each survey had its own, unknown effort, which was considered constant throughout the study region and was given a log-Gaussian prior distribution. Other examples of analyses with non-constant survey effort are presented by Renner and Warton (2013) and Warton et al. (2013), who consider modeling presence only data. In their applications, the effort is a function of environmental covariates and spatial location, $z(s,x)$. Yuan et al. (2017) consider LGCP models for transect surveys where the observation probability decreases with distance from survey vessel.

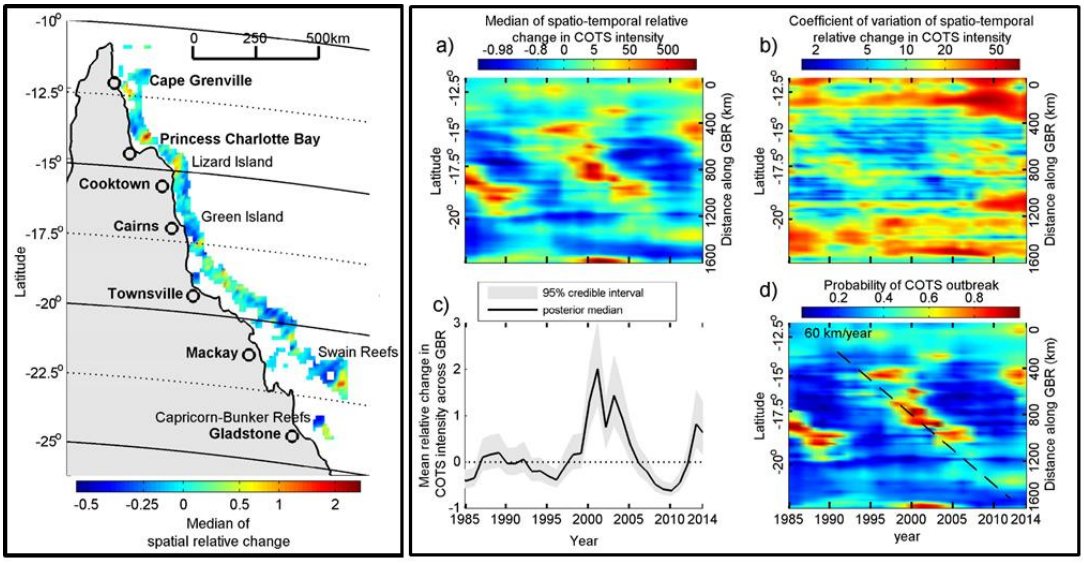
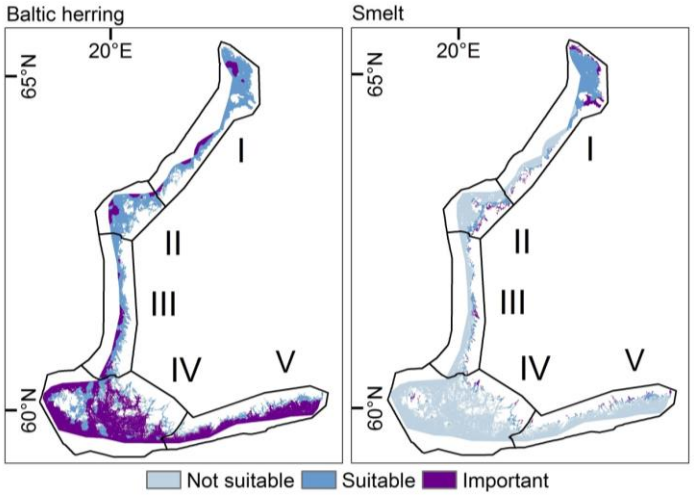


Figure 2.2.2. Illustration of spatial (left) and spatial-temporal (right) random effects reported as relative change/effect in intensity in the Great Barrier Reef, Australia. The spatial-temporal random effect is projected on the reef system's center line. In plot c) the outbreak corresponds to a 2-fold increase in Crown of Thorns starfish abundance relative to the local mean abundance. Reproduced with permission from (Vanhatalo et al., 2017).



	Total number of larvae in surface water, ×10⁹ mean (95% credible interval)	Percent of water area suitable for larvae	Percent o water are producin 80% of larvae
perch	1.56 (0.89, 2.55)	13.66	3.03
pikeperch	0.54 (0.12, 1.56)	3.68	1.37
Baltic herring	8.72 (5.65, 12.86)	99.79	52.89
smelt	5.91 (2.88, 10.81)	22.50	4.44

Figure 2.2.3. Illustration of abundance predictions and classification of spatial regions based on their relative importance to fish production. Reproduced with permission from (Kallasvuo et al., 2017).

Joint species distribution models

Joint species distribution models (JSDM) have gained increasing interest in recent years (Warton et al., 2015). The key assumption behind most of the JSDMs is that the model should include interspecific dependence between species-specific environmental effects and spatial random effects. In our context, extension to JSDMs corresponds to denoting the species-specific log intensity with:

$$f_j(s, x_t) = \alpha_j + x_s{}^T \beta_j + \varphi_j(s),$$

where j denotes a species and gives the species-specific parameters a hierarchical prior/model that introduces the dependence. For example, in the hierarchical model of species communities (HMSC, Ovaskainen et al., 2017), the species-specific linear weights are given a Gaussian prior $\beta_j \sim N(\mu_j, V)$, where μ_j is the mean vector and V the covariance matrix. The interspecific dependence is then included by, for example, modeling the mean vector as a function of species traits; that is $\mu_j = \sum t_{jl}\tau_{lk}$ where t_{jl} is the l'th trait of the species and τ_{lk} are the trait weights. The HMSC allows also for a generic multivariate Gaussian prior for $\mu_1, \ldots \mu_J$, as well as inclusion of phylogenetic information into the correlation between $\beta_1, \ldots \beta_J$. In species archetype modeling the responses of the species to the environment are clustered into a few archetype models corresponding to shared β between group of species (Dunstan et al., 2013; Hui et al., 2013). Similarly, the spatial random effects are extended to include interspecific correlations. Vanhatalo et al. (manuscript) extended point process based JSDMs to semiparametric models, where the effects of environmental covariates are allowed to follow a semiparametric Gaussian process model.

Key advantages of JSDMs compared to single species models can be summarized as follows. The inclusion of interspecific dependence allows information flow between species, which improves the estimates for covariate effects especially for species with only scarce data (e.g., Ovaskainen et al., 2017; Hui et al., 2013; Clark et al., 2014). From the predictive point of view, this approach has added benefit that the models predictive accuracy also improves in both interpolation (Ovaskainen et al., 2017; Hui et al., 2013) and extrapolation (Vanhatalo et al., manuscript). JSDMs have also been demonstrated to more accurately estimate the joint distribution of multiple species, whereas single species models typically predict larger distribution areas than JSDMs (see Figure 2.2.4). Moreover, the distribution of species is not governed only by their environment and stochastic processes and species interactions play a significant role in the realized distribution (Ovaskainen et al., 2017). The interspecific correlations in the spatial random effects account also for the species-to-species interactions. By examining the interspecific correlation matrices, we can gain understanding of the similarities in the response of species to the environment, and of species-to-species interactions (see Figure 2.2.5 for illustration).

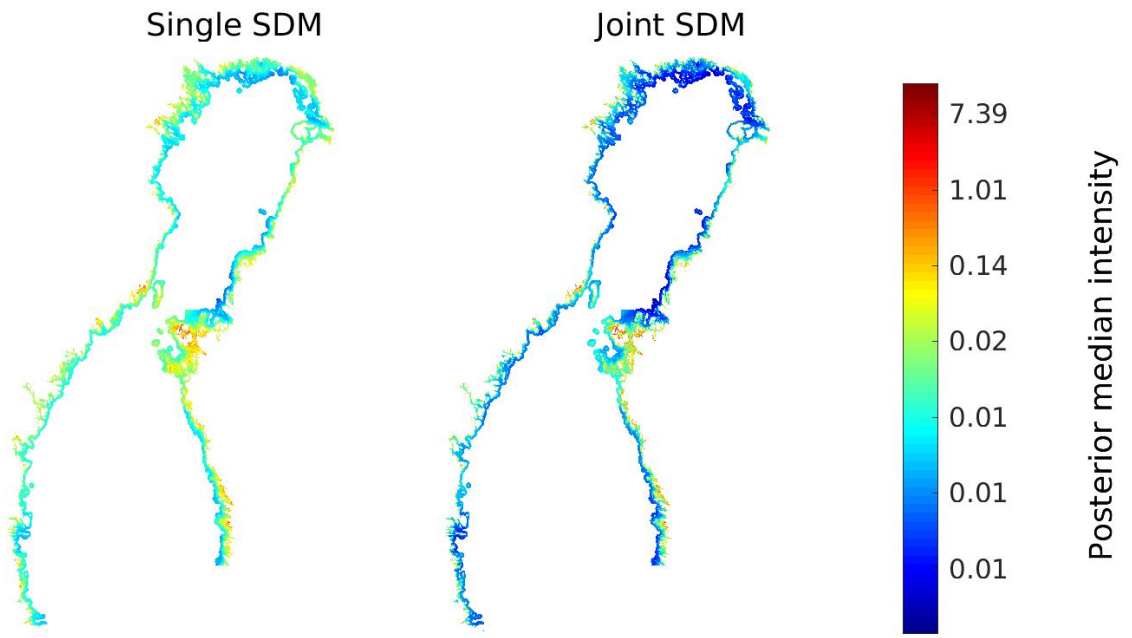


Figure 2.2.4. Abundance predictions with single and joint species distribution models for the three-spined stickleback in the Gulf of Bothnia, northern Baltic Sea. The JSDM typically predicts smaller distribution ranges. Reproduced with permission from (Vanhatalo et al., manuscript).

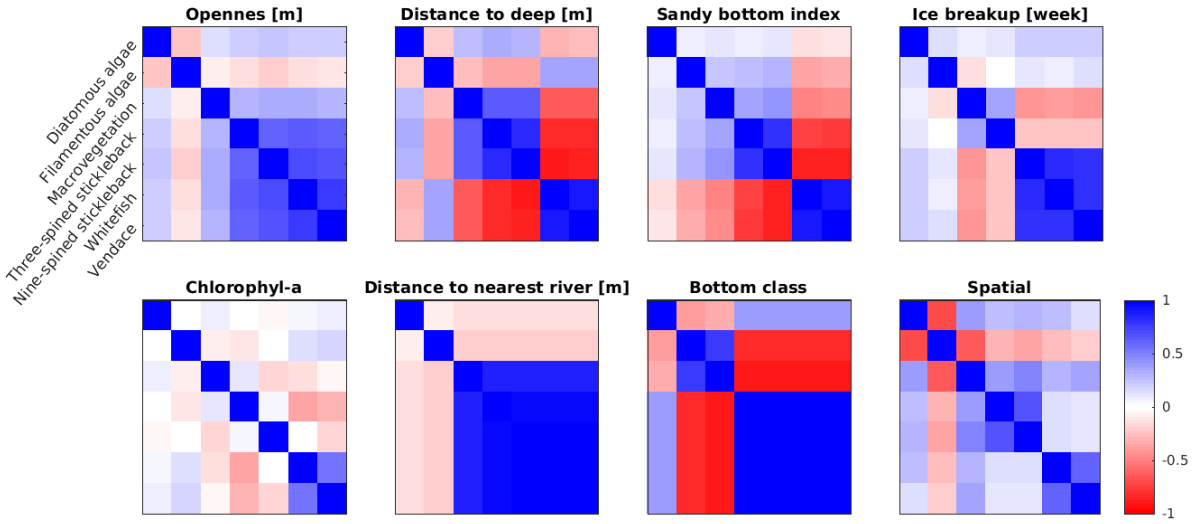


Figure 2.2.5. Interspecific correlations between the fixed effects, β_j, (first 7 plots) and spatial random effect. Reproduced with permission from (Vanhatalo et al., manuscript).

Discussion

Point process models are classical tools in spatial statistics as well as in some areas of applied ecology. In recent years, they have gained increasing attention in the general species distribution modeling literature as well. In this working paper, I discussed their connection to traditional generalized linear model-based species distribution models, as well as to current developments in joint species distribution models. I also highlighted some of their useful properties from a practical modeling viewpoint. Namely, the point process framework provides results that are directly interpretable as (relative) species densities.

The case studies reviewed in this work included classical count data observed either visually or collected with fishing nets. However, an interesting area for future development would be to extend the methods to other types of data as well. In the deep-sea context this could include, for example, combining trawling and visual data with acoustic survey data. Juntunen et al. (2012) proposed a Bayesian method to integrate trawling and acoustic survey data to estimate species composition and abundance (biomass) of multiple fish in the Baltic Sea. Their approach does not follow the point process framework as such, but is very closely related to it, and could in principle be extended to it as well.

References

Banerjee, S.Carlin, B. P., and Gelfand, A. E. 2015. Hierarchical Modeling and Analysis for Spatial Data. Second Edition. CRC Press/Chapman & Hall. Monographs on Statistics and Applied Probability 135, Boca Raton, Florida, 562 pp.

Clark, J.S., Gelfand, A.E., Woodall, C.W. and Zhu, K. 2014. More than the sum of the parts: forest climate response from joint distribution models. Ecol. Appl. 24: 990-999.

Dunstan, P. K., Foster, S. D., Warton, D. I. and Hui, F. K.C. 2013. Finite mixture of regression modeling for high-dimensional count and biomass data in ecology. J. Agric. Biol. Environ. Stat. 18: 357-375.

Hui, G.K.C., Warton, D.I., Foster, S.D. and Dunstan, P.K. 2013. To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. Ecol. 94: 1913-1919.

Juntunen, T., Vanhatalo, J., Peltonen, H., and Mantyniemi, S. 2011. Bayesian spatial multispecies modeling to assess pelagic fish stocks from acoustic- and trawl-survey data. ICES J. Mar. Sci. 69: 95–104. https://doi.org/10.1093/icesjms/fsr183

Kallasvuo, M., Vanhatalo, J., and Veneranta, L. 2017. Modeling the spatial distribution of larval fish abundance provides essential information for management. Can. J. Fish. Aquat. Sci. 74: 636-649.

Latimer, A., Wu, Shanshan, Gelfand, A.E. and Silander, J. 2006. Building statistical models to analyze species distributions. Ecol. Appl. 16: 33-50.

Mäkinen, J. A.-E., and Vanhatalo, J.P. 2018. Hierarchical Bayesian model reveals the distributiona shifts of Arctic marine mammals. Divers. Distrib. 2018: 1-14; https://doi.org/10.1111/ddi.12776

Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., Roslin, T. and Abrego, N. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett. 20: 561-576.

Renner, I.W., and Warton, D.I. 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. Biometrics 69: 274-281.

Vanhatalo, J., Veneranta, L., and Hudd, R. 2012. Species distribution modeling with Gaussian processes: a case study with the youngest stages of sea spawning whitefish (*Coregonus lavaretus* L. s.l.) larvae. Ecol. Model. 228: 49-58.

Vanhatalo, J., Hosack, G., and Sweatman, H. 2017. Spatio-temporal progression of outbreaks of the crown-ofthorns starfish on the Great Barrier Reef, 1985–2014. J. Appl. Ecol.54: 188-197.

Vanhatalo J., Hartmann, M. and Veneranta, L. (manuscript). Joint species distribution modeling with additive multivariate Gaussian process priors and heteregenous data. arXiv:1809.02432v1

Warton, D.I., Blanchet, F.G., O-Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., and Hui, F.K.C. 2015. So many variables: joint modeling in community ecology. Trends Ecol. Evol. 30: 766-779.

Warton, D. I., Renner, I. W., and Ramp, D. 2013. Model-based control of observer bias for the analysis of presence-only data in ecology. PLoS ONE, 8(11). https://doi.org/10.1371/journal.pone.0079168

Yuan, Y., Bachl, F. E., Lindgren, F., Brochers, D. L., Illian, J. B., Buckland, S. T., Ruie, H., and Gerrodette, T. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann. Appl. Stat. 11(4), 2270–2297. https://doi.org/10.1214/16-AOAS1011

2.2.1. Group Discussion

The discussion following Dr. Vanhatalo's presentation examined issues surrounding specifics of species distribution modeling and joint species distribution modeling (JSDM), as well as exploring the user-friendliness of various models, and how this attribute affects model usefulness. Dr. Vanhatalo posited that the point process approach described in his presentation was in essence no

more complicated than a standard GLM approach in R, and felt that any user who could use the latter would have no difficulty in implementing the former. An instructional/guidance document complete with sample data and worked examples will be made publicly available to improve the ease with which interested parties might adopt the point process approach described in Dr. Vanhatalo's talk.

Point process models can be very useful where a user wants to make use of data with high degrees of uncertainty to improve modeling performance. For example, data from acoustic monitoring equipment can be helpful in detecting how many distinct species of fish are in a given area (however, further data would be required to try to identify those species etc.). Observational data (e.g., data from acoustic monitoring) has quantitative aspects, which can aid in analysis, and the amount of information with these data reduces the residual variance because more aspects of these data are explained. The discussion also focused on how the MaxEnt (maximum entropy) model approach compares with the spatial point process. The point was raised that many scientists have short timeframes to produce results and MaxEnt is a very user-friendly option of model. Other, more complicated models are being developed but are not as easy to use. The discussion indicated that, if MaxEnt is used properly, it can return excellent results and can produce a similarly smart analysis. However, there are benefits to investing time in learning more about model-based approaches because in the long term they will likely give more information alongside the model output. Following this discussion of MaxEnt's user-friendly nature, the point was raised that no matter which type of model is being used, it is still being written in R as a type of Generalized Linear Model (GLM) at the base level. The challenge arrives once the user goes beyond R packages and focuses on more complex data types and spatial correlations. It is important to consider the analysis before fieldwork is completed because often datasets do not correspond well with the analysis that will be done. One method of helping users become more familiar with these methods is producing test datasets alongside publications, which can be used as trial projects for users and gives them the tools to do such analyses properly. It was agreed that the provision of trial datasets is an important aspect of increasing the uptake of new complex methods.

3. EXTENDED ABSTRACTS

Participants were invited to provide extended abstracts for this publication on a topic of their choice. Some authors volunteered to present their subject more formally in the Participants Forum on Day 1 (Appendix 1), and the topic was discussed further by the group. For those abstracts a summary of the group discussion follows the extended abstract. The formatting follows the original style of the submission, and figures are numbered consecutively within each abstract with the numbering restarting with the next.

3.1. What do Environmental Managers and Stakeholders want from a Model?

Ashley A. Rowden

National Institute of Water & Atmospheric Research, Wellington, New Zealand

Human activities are increasingly impacting the deep-sea environment, and there is a need to manage the effects of these impacts on deep-sea species. To design effective spatial management measures to conserve and protect deep-sea species it is necessary to know their distribution patterns. Environmental managers and stakeholders understand that it is necessary to use numerical models to determine distribution patterns because of the paucity of deep-sea species records. However, the acceptance and use of species distribution models in spatial management planning can depend on several factors that model-building scientists don't always perceive or address. Recent experience producing species distribution models for Vulnerable Marine Ecosystems (VMEs), as part of a process to design management measures for deep-sea bottom trawling in the South Pacific, has highlighted what environmental managers and stakeholders want from a model. These need to include the following: A model should to be relevant; models for species that act as proxies or indicators for VMEs are not as acceptable as models that attempt, even coarsely, to model the ecosystem feature itself (e.g., coral reef). A model should be made at the appropriate spatial scale; because it is possible, models are sometimes made at a resolution that is smaller or bigger than the use to which they will be put, but this impacts upon their usefulness and acceptance (e.g., making a model at 1 km^2 resolution when spatial closures will be applied at a scale of >1000 km^2). The modeling approach should be conservative and parsimonious; multiple model or ensemble approaches are preferred over reliance on a single model approach (even if considered by scientists to be the latest and best), and models using fewer predictor variables are more easily understood and accepted. A model should to be believable; field validation of a model is highly valued (even for a small portion of the modelled area), and the more broadly understood metrics used for internal model validation are preferred over those that are more model-approach specific (e.g., correlation metrics versus AUC). A model's limitations must be expressed spatially; quantifying and mapping the spatial uncertainty of a model helps acceptance (acknowledgment that model not equally good everywhere), and is a particularly useful output for designing spatial management measures. In the main, these needs are similar or the same as the concerns of a model-building and evaluating scientist, but there are subtle differences in viewpoint that impact acceptance of models by managers and stakeholders. This presentation will expand upon these differences with the aim of assisting in the making of models and their outputs which are more readily accepted and used by environmental managers and stakeholders for spatial management planning.

3.1.1. Group Discussion

The discussion following this talk focused on concerns arising from designating vulnerable marine ecosystems (VMEs) and which thresholds should be used to quantify these boundaries (e.g.,

where a VME indicator species is at a "significant concentration" that represents a VME). A comment was raised questioning how best to move from a species approach to a VME approach and whether the threshold of abundance of a VME indicator species should be used.

A second comment questioned the possibility of jointly modeling species and traits, and whether this can be used to designate VMEs. This approach would be useful because the FAO Guidelines for the identification of VME indicators provides a list of traits that such species should possess.

Finally, it was suggested that model predictions for species habitat should (when published) be compared to areas where stakeholders and managers want to ban fishing. These maps can often show different areas, and for this reason, communication between model creators, managers, and stakeholders has to be increased to ensure that management decisions are properly informed by the best available information.

3.2. Rough Data: Imperfect Data and Directional Rugosity

Cherisse Du Preez, Emily Rubidge and Jessica Finney

Pacific Region, Fisheries and Oceans Canada (DFO)

This working paper covers two topics applicable to species distribution models (SDMs) in the deep sea: imperfect data and seafloor roughness (rugosity). The following is based on the experiences of Fisheries and Oceans Canada scientists who modeled vulnerable marine ecosystems and ecologically or biologically significant areas in the deep Northeast Pacific.

Imperfect data and transparency about it

Imperfect detection

The probability of detecting a species at a site when it is present is calculated through repeated sampling of the same site, and combined with statistical models to estimate species' occupancy. Although factoring in the probability of detection is standard in terrestrial ecology, it has proven impractical in the deep sea. Why? Cost, time, accessibility… If you haven't already, may I suggest reading "How long should we ignore imperfect detection of species in the marine environment when modeling their distribution?" by Jacquomo Monk (2014). In response to ideas like those discussed by Monk, a team of DFO researchers are planning a repeat survey study using remotely operated vehicles and visual benthic surveys.

Uncertainty in extrapolations vs. interpolations

It is common practice to model beyond the range of data sampled for one or more environmental variables, be it carefully and within reason. A shaded overlay on an uncertainty map is often used to identify such extrapolations (Figure 1). An extrapolated area analyses is a standard SDM output,

but there isn't an equivalent for interpolation, to identify data gaps within a range sampled. During the 2012 Cobb Seamount expedition, the team sampled between 34 and 1154 m depth, but technical issues prevented them from surveying between 211 and 472 m depth. The ~250 m gap occurred over problematic depths, at the transition region from flat plateau to steep flanks, and the known narrow depth ranges of unique cold-water coral gardens and reefs and dense lost-fishing gear (discussed in Du Preez et al., 2016). To clearly identify the unreliable interpolation within the range of data sampled, the team manually added a hatching overlay to their uncertainty map (Figure 1).

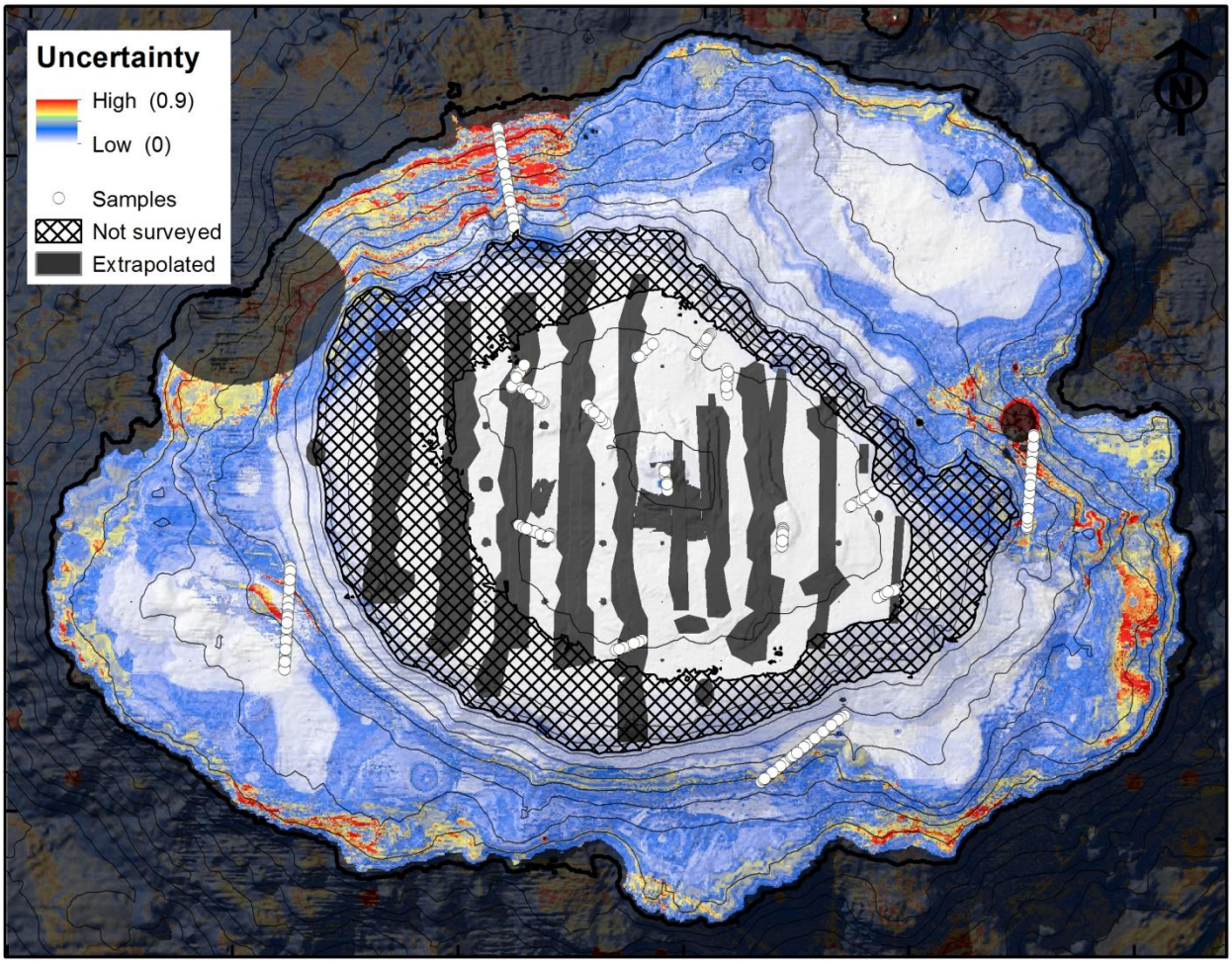


Figure 1. Outputs from the Random Forest model on Cobb Seamount illustrating the extrapolated areas (grey shading) and the unreliable interpolated area (i.e., 211 to 472 m depth gap not surveyed; hatching) (Du Preez et al., 2016).

Combination models

If you are fortunate enough to have density input data for modeling, consider the value of applying a *presence probability* threshold to your *density predictions* (e.g., only display predicted density if output presence probability is > 0.5). While the drawback is that densities will not be displayed

over the full study area, the advantage is that they will only be displayed where there is a higher probability that the species is present. A similar result can be achieved by applying an *uncertainty* threshold (a common species distribution modeling output).

Directional rugosity

Seafloor roughness (or complexity) is a common environmental variable used in species distribution models. There are many metrics for measuring seafloor roughness. One useful metric is the arc-chord ratio (ACR) (Du Preez, 2015), which is part of the ESRI ArcGIS Benthic Terrain Modeller toolbox (Walbridge et al., 2018). Regardless of the specific roughness metric used, scale and directionality are two important factors to consider. The former is a fairly common consideration, while the latter is often overlooked, especially when the 3-D bathymetric geoprocessing uses a moving window or circular polygon. In environments with dominant flow directionality (uni- or bi-directional), using a linear metric of rugosity upstream may provide a strong indicator for species distributions and other biological characteristics (e.g., biodiversity) downstream (Figure 2). Physical mechanisms may include changes in pressure, dissolved gases, shear stress, particulate load, entrained organisms and larvae, and water stratification. Changes in the local bottom flow regime (e.g., changes to flow velocity, energy, and directionality) affect epibenthic communities through larval delivery and recruitment, delivery of oxygen and nutrients, feeding opportunities, removal of waste, the passive collection or dispersal of organisms, suspension and deposition of sediment (i.e., available substratum and turbidity), scouring and erosion of sediment and of organisms, and levels of biotic and abiotic disturbance (likely influencing sessile fauna more than mobile fauna). Incorporating directionality into roughness metrics may improve the predictive power of species distribution models over conventional rugosity metrics. Directional linear ACR rugosity will be released within the ESRI ArcGIS Benthic Terrain Modeller soon.

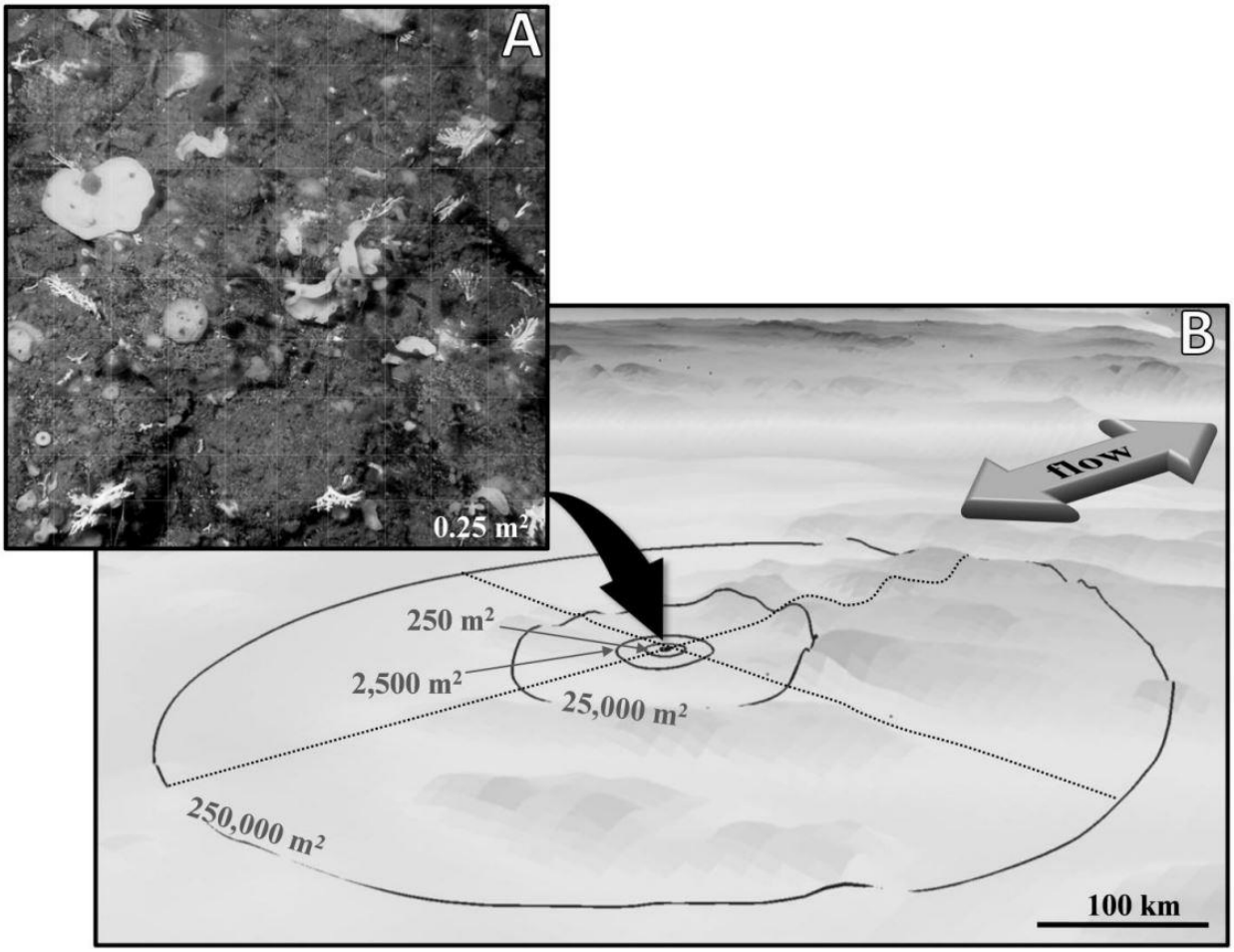


Figure 2. In a study comparing different scales, dimensions, and directionality, large-scale linear rugosity upstream was found to be the best predictor of local benthic biodiversity (Du Preez, 2014).

References

Du Preez, C. 2014. Resolving relationships between deep-sea benthic diversity and multi-scale topographic heterogeneity (Doctoral dissertation, Biology Dept., University of Victoria, British Columbia, Canada). https://dspace.library.uvic.ca/handle/1828/5828?show=full

Du Preez, C. 2015. A new arc–chord ratio (ACR) rugosity index for quantifying three-dimensional landscape structural complexity. Landscape Ecol. 30: 181-192.

Du Preez, C., Curtis, J.M. and Clarke, M.E. 2016. The structure and distribution of benthic communities on a shallow seamount (Cobb Seamount, Northeast Pacific Ocean). PloS One 11(10), p.e0165513.

Monk, J. 2014. How long should we ignore imperfect detection of species in the marine environment when modeling their distribution? Fish Fish.15: 352-358.

Walbridge, S., Slocum, N., Pobuda, M. and Wright, D.J. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler. Geosci. 8: 94.

3.3. Determining the best methods for model validation

Chris Rooper*, Rachel Wilborn and Pam Goddard

Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA

*Current affiliation: Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

Model validation for species distribution modeling usually takes one of three forms; resubstitution or resampling, testing with new or held back data, or testing with independent data collected from a different area or an independent survey. The general purpose of model validation exercises is most often to determine the robustness of the model and the confidence that can be placed in a model's predictions.

Within-sample model validation (resubstitution or resampling) is the most commonly used method for model evaluation, typically as a leave-one-out analysis or a k-fold cross-validation, where the effects of different partitions of the modeled data on the results are examined. Testing with held back data from the same modeled area (either from a new time or a random selection of data) is also fairly common. Transferring the model to a new area or a newly collected data set designed to test model predictions are fairly uncommon methods, possibly because of the associated high cost of additional sampling. The literature (and practical experience) indicates that predicting an entirely new data set from a new area is probably the most difficult test of any species distribution model.

Models are usually evaluated using the same criteria as the initial models (e.g., using AUC, TSS and R^2). An example of a model validation is shown in Figure 1. These results are from SDMs of structure-forming invertebrates in the Aleutian Islands of Alaska. Predictions of species distributions were initially developed using Generalized Additive Models and bottom trawl survey data. The models were tested on a year of data that was held back (2012). The results of this testing were good for both presence-absence models and abundance models (with the exception of Stylasteridae).

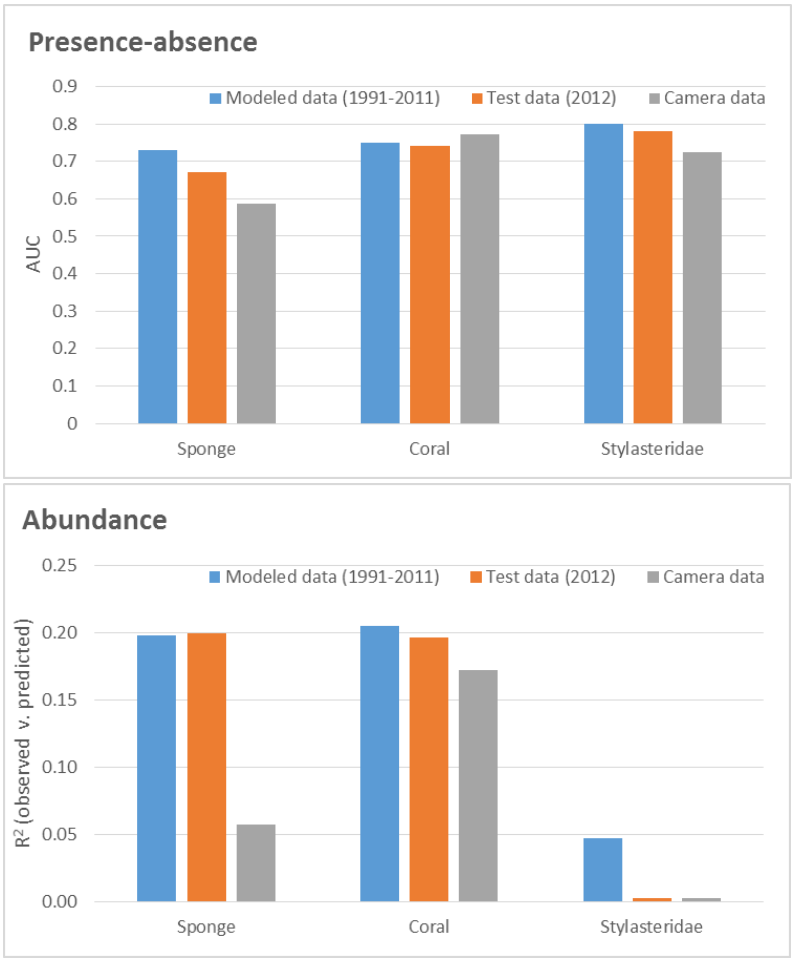


Figure 1. AUC values and R^2 values for coral and sponge models developed using bottom trawl survey data in the Aleutian Islands. Test data was a year (2012) that was held back for testing purposes and camera data was an independent camera survey completed in 2014. Details of these results can be found in Rooper et al. (2014) and Rooper et al. (2018). Sponges are combined classes (Hexactinellida and Demospongiae) and corals include the order Antipatharia, suborders Holaxonia (families Plexauridae, Acanthogorgiidae), Calcaxonia (families Primnoidae and Isididae), Scleraxonia (family Paragorgiidae), family Paramuriceidae, and hydrocorals from the family Stylasteridae.

In 2012 and 2014 an independent validation survey was conducted at 216 randomly selected sites in the Aleutian Islands. These data used a different gear (underwater camera) and surveyed a somewhat different area, as camera transects were able to be conducted in rough, rocky areas where the bottom trawl could not be deployed. The results were still adequate for presence absence and abundance models of coral. However, the ability of the bottom trawl survey models to predict sponge presence or absence and abundance in the camera survey was not good. This result demonstrates that challenging a species distribution model with new and different data can

sometimes reduce the confidence in its performance relative to the performance on resampled or held back data.

Some potential questions for discussion are:

- What are the best methods for testing species distribution models (especially when collecting new data is not feasible)
- Are there better ways to evaluate model performance in validation exercises?
- How good does model performance in validation exercises need to be in order to accept a model?

References

Rooper, C.N., Wilborn, R.E., Goddard, P., Williams, K., Towler, R., and Hoff, G.R. 2018. Validation of deep-sea coral and sponge distribution models in the Aleutian Islands, Alaska. ICES J. Mar. Sci. 75:199-209.

Rooper, C.N., Zimmermann, M., Prescott, M., and Hermann, A. 2014. Predictive models of coral and sponge distribution, abundance and diversity in bottom trawl surveys of the Aleutian Islands, Alaska. Mar. Ecol. Prog. Ser. 503:157-176.

3.3.1. Group Discussion

Dr. Rooper's presentation explored three main validation techniques: resubstitution or resampling, testing with new or held back data, or testing with independent data collected from a different area or independent survey. The discussion following this presentation focused mainly on bias inherent with different types of data collection and model validation. To begin, it was noted that one must use the same method of data collection to both test and train the data (for example camera data, trawl data, etc.). Next, trawl data, ROV (Remotely Operated Vehicle) data, drop cameras and other methods of data collection all come with biases in terms of the area they describe. Validating models with ROV data is problematic due to the data often being comprised of a 500 m by 500 m cell, while the ROV moves only through the middle of this area, therefore missing large amounts of seafloor. Validating models with drop camera data was mentioned as resulting in less bias for habitat mapping than ROVs due to their blind sampling, however, drop cameras stop working well on steep slopes and potentially in other areas such as very soft sediments where plumes can be created. This limitation may result in a similar bias if such areas are included in the study area, because it limits possible deployment areas.

3.4. Distribution Models Applied to Climate Change in the Deep Sea. A Promising but Challenging Development Field

José Manuel González-Irusta and Telmo Morato

Departamento de Oceanografia e Pescas, Universidade dos Açores
PT-9901-862 HORTA, Portugal

Species distribution models are being currently used to inform marine spatial planning worldwide, modeling the distribution of relevant species (González-Irusta et al., 2015; Greathead et al., 2015; Parra et al., 2016), the distribution of benthic assemblages as a proxy to biological habitats (Serrano et al., 2017) or the abundance of structural species such as *Lophelia pertusa* as a proxy to the habitat itself (Howell et al., 2011; Fernandes et al., unpublished). These models have also been successfully used to delineate essential fish habitats of commercial fish species, modeling the abundance of juveniles (nursery areas, Aires et al., 2014; Asjes et al., 2016), the abundance of adults at their spawning stage (spawning grounds, González-Irusta and Wright, 2016a, 2016b, 2017) and fish egg distribution (e.g., Loots et al., 2011; Lelièvre et al., 2014). More recently, these models have been combined with other type of approaches to answer different questions than species presence or habitat location. For instance, distributions have been combined with Biological Traits Analysis and effort maps (from Vessel Monitoring Systems) to determine and map the impact of trawling disturbance on sensitive species (González-Irusta et al., 2018). They have also been combined with ecosystem simulation models to produce maps of ecosystem production across entire ecosystems such as the Gulf of Mexico (Grüss et al., 2018), and their use combined with particle tracking analysis in connectivity studies has been already proposed (e.g., Gallego et al., 2016). Finally, these models have been combined with future climatic scenarios to predict distribution changes for different marine species (e.g., Tittensor et al., 2010; Sequeira et al., 2014; Gallego et al., 2017). The combination of species distribution models with other techniques (such as biological traits analysis or community analysis), and its use to predict climate change impacts on marine biodiversity, are probably two of the most promising fields for the implementation of the ecosystem approach into the management of the marine ecosystems. In the framework of the ATLAS project, we are currently working on the use of these models to predict changes in future distributions of cold-water coral species and fish under different climate change scenarios at a North Atlantic basin scale. Although some of the challenges we are facing are common to most of the model exercises (lack of biological data, sampling bias, poor resolution in some environmental layers, etc.) others are specific to the application of these models to predict future changes in distribution. One of the challenges is the environmental variable selection and its impact on the models' sensitivity to future changes. One example of this issue, especially important for the deep sea, is the inclusion or not of depth into the models. In fact, depth is not an environmental variable *stricto sensu* but usually is the variable with the most accurate information, and a very good proxy for several other relevant variables (such as temperature near bottom, salinity, light intensity, pressure, etc). To include depth in the models will improve them in most

cases, but can also reduce the sensitivity of these models to future changes in environmental conditions. A better understanding of the relationship between the modelled species and depth is a key factor to make the right decision. Another important challenge is the lack of appropriate tools to predict the accuracy of these models to extrapolate their present prediction into the future. Current evaluation methods such as AUC and TSS are controversial even for present predictions (Lobo et al., 2008; Jiménez-Valverde, 2014). Fourcade et al. (2018) has recently demonstrated the inability of current evaluation metrics to assess the biological significance of the relationship between the response variable and the explanatory variables in distribution models, if they are not based on a solid knowledge of the species and its ecological niche. This issue is especially relevant when these models are used to extrapolate present predictions into the future based on these relationships. Species distribution models can predict accurately the distribution of habitat and species even based on spurious relationships, but they will only be able to predict future changes if the model is really based on the ecological niche of the species. A sound knowledge of the environmental requirements of the species, the support of experimental work on this topic, and the inclusion of the whole range of environmental variability for the target species are key for producing a correct distribution model that is able to predict future changes in species distribution under different climate change scenarios.

References

Aires, C., González-Irusta, J.M., and Watret., R. 2014. Updating Fisheries Sensitivity Maps in British Waters. Scottish Marine and Freshwater Science Vol 5 No 10. 88 pp. DOI: 10.7489/1555-1.

Asjes, A., González-Irusta, J. M., and Wright, P. J. 2016. Age-related and seasonal changes in haddock *Melanogrammus aeglefinus* distribution: Implications for spatial management. Mar. Ecol. Progr. Ser. 553: 203-217.

Fernandes, P. G., McIntyre, F.D., González-Irusta, J.M., and Neat, F.C. (2018). Estimating the abundance of deep cold-water coral: the Northeast Atlantic's great barrier reef. *Unpublished*

Fourcade, Y., Besnard, A.G., and Secondi, J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecol. Biogeogr 27: 245-256.

Gallego, A., Gibb, F. M., Tullet, D., and Wright, P. J. 2016. Bio-physical connectivity patterns of benthic marine species used in the designation of Scottish nature conservation marine protected areas. ICES J. Mar. Sci. 74: 1797–1811.

Gallego, R., Dennis, T. E., Basher, Z., Lavery, S., and Sewell, M. A. 2017. On the need to consider multiphasic sensitivity of marine organisms to climate change: a case study of the Antarctic acorn barnacle. J. Biogeography 44: 2165–2175.

González-Irusta, J. M., González-Porto, M., Sarralde, R., Arrese, B., Almón, B., and Martín-Sosa, P. 2015. Comparing species distribution models: A case study of four deep sea urchin species. Hydrobiologia 745: 43–57.

González-Irusta, J. M., and Wright, P. J. 2016a. Spawning grounds of Atlantic cod (Gadus morhua) in the North Sea. ICES J. Mar. Sci. 73: 304–315.

González-Irusta, J. M., and Wright, P. J. 2016b. Spawning grounds of haddock (Melanogrammus aeglefinus) in the North Sea and West of Scotland. Fish. Res., 183: 180–191.

González-Irusta, J. M., and Wright, P. J. 2017. Spawning grounds of whiting (Merlangius merlangus). Fish Res. 195: 141–151.

González-Irusta, J. M., De la Torriente, A., Punzón, A., Blanco, M., and Serrano, A. 2018. Determining and mapping species sensitivity to trawling impacts: the BEnthos Sensitivity Index to Trawling Operations (BESITO). ICES J. Mar. Sci. https://doi.org/10.1093/icesjms/fsy030

Greathead, C., González-Irusta, J. M., Clarke, J., Boulcott, P., Blackadder, L., Weetman, A., and Wright, P. J. 2015. Environmental requirements for three sea pen species: relevance to distribution and conservation. ICES J. Mar. Sci. 72: 576–586.

Grüss, A., Drexler, M. D., Ainsworth, C. H., Babcock, E. A., Tarnecki, J. H., and Love, M. S. 2018. Producing Distribution Maps for a Spatially-Explicit Ecosystem Model Using Large Monitoring and Environmental Databases and a Combination of Interpolation and Extrapolation. Front. Mar. Sci. 5: 1–20.

Howell, K. L., Holt, R., Endrino, I. P., and Stewart, H. 2011. When the species is also a habitat: Comparing the predictively modelled distributions of *Lophelia pertusa* and the reef habitat it forms. Biol. Cons. 144: 2656–2665.

Jiménez-Valverde, A. 2014. Threshold-dependence as a desirable attribute for discrimination assessment: implications for the evaluation of species distribution models. Biodivers. Conserv. 23: 369–385.

Lelièvre, S., Vaz, S., Martin, C. S., and Loots, C. 2014. Delineating recurrent fish spawning habitats in the North Sea. J. Sea Res. 91: 1–14.

Lobo, J. M., Jiménez-Valverde, A., and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17: 145–151.

Loots, C., Vaz, S., Planque, B., and Koubbi, P. 2011. Understanding what controls the spawning distribution of North Sea whiting (Merlangius merlangus) using a multi-model approach. Fish. Oceanogr. 20: 18-31.

Parra, H. E., Pham, C. K., Menezes, G. M., Rosa, A., Tempera, F., and Morato, T. 2016. Predictive modeling of deep-sea fish distribution in the Azores. Deep Sea Res. Part 2. Top. Stud. Oceanogr. 145: 49-60.

Sequeira, A. M. M., Mellin, C., Fordham, D. A., Meekan, M. G., and Bradshaw, C. J. A. 2014. Predicting current and future global distributions of whale sharks. Global Change Biol. 20: 778–789.

Serrano, A., González-Irusta, J. M., Punzón, A., García-Alegre, A., Lourido, A., Ríos, P., Blanco, M., *et al.* 2017. Deep-sea benthic habitats modeling and mapping in a NE Atlantic seamount (Galicia Bank). Deep Sea Res. Part 1. Oceanogr. Res. Pap. 126: 115–127. http://dx.doi.org/10.1016/j.dsr.2017.06.003.

Tittensor, D. P., Baco, A. R., Hall-Spencer, J. M., Orr, J. C., and Rogers, A. D. 2010. Seamounts as refugia from ocean acidification for cold-water stony corals. Mar. Ecol. 31: 212–225.

3.4.1. Group Discussion

To begin the discussion, the concept of evolutionary adaptability was introduced. It was noted that most correlative models assume to capture the process of evolutionary adaptability, unlike mechanistic models. A question was raised as to whether there is a way to incorporate the mechanistic component into statistical models and not completely rely on correlative models. Mechanistic distribution models look at the physiological attributes of an animal and can determine factors needed for survival (i.e., water, food, etc.), and potentially suitable areas can then be determined based on variable thresholds (a fundamental niche). However, there are often insufficient data on the physiological attributes of deep-sea species. It was also noted that it is important to consider for what use future predictions are being developed; many scientists extrapolate to 2050, 2100 and dates such as this, but managers and stakeholders are likely to be working on a shorter time frame, generally decadal, so what would a 2100 prediction be used for? If one wanted to look specifically at smaller sections of one larger area, it may be beneficial to look at areas where one knows conditions will change (i.e., salinity, temperature), so perhaps only these regions could be used to apply species distribution models. Next, it was mentioned that as a field of study, we need a greater experimental understanding of how animals respond to changes in specific variables; many animals can still survive certain levels of change and more experimental tests would be helpful in order for models to be more applicable. Finally, it was mentioned that many studies that model the future distribution of VME indicator taxa do not take into account reproduction and conditions necessary for reproduction, and rather focus on mature animals. This focus has led to a large gap in knowledge and possibly unrealistic predictions.

3.5. Including Predictions of Community Functional Traits in Species Distribution Models using Hierarchical Modeling of Species Communities (HMSC)

Benjamin Weigel

Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki 00014, Finland

Background

The ever-increasing impact of human-induced pressures on marine ecosystems causes reorganization of communities by reshuffling species compositions, altering distribution ranges and affecting species interactions, through changing environmental parameters and physical disturbances. Geographic distribution, as well as species interactions within a community, ultimately depend on the individual traits of a species. Functional traits may reflect information on the species' life history strategies, as well as morphological, physiological and behavioural characteristics. Such information provides insight into, for example, preferred habitat types, feeding modes or size distribution, among various other aspects that can be linked to ecosystem functions and services (Weigel et al., 2016). Hence, considering species traits in distribution modeling is a promising tool to better understand and predict community developments under environmental changes, while also including the functional implications of community changes, and therefore may be a useful framework for management and conservation.

Aim and Approach

The aim of this working paper is to highlight the possible application of including functional trait information within Hierarchical Modeling of Species Communities (HMSC; Ovaskainen et al., 2017), an approach that belongs to the class of joint species distribution models (JSDM; Warton et al., 2015). To illustrate some of the possibilities HMSC allows for, I used zoobenthos community datasets extracted from the EMODnet data base (www.emodnet-biology.eu) for the Gulf of Finland (Finnish Environment Institute SYKE; 2018), including a total of 93 species at 541 locations that have been sampled between 2009 and 2014. I used time and depth as explanatory variables. The trait data was obtained from the MERP Trait Explorer (http://www.marine-ecosystems.org.uk/Trait_Explorer, Bruggeman et al., 2009; Brey et al., 2010; Webb and Hosegood, 2013; Webb et al., 2017) offering inferred trait information for all marine species, while indicating the uncertainty of their values based on available data. I included traits such as maximum size and maximum mass of the species (or family) as well as feeding types, to illustrate this approach.

Preliminary Results and Conclusions

In Figure 1, I highlight some relevant outputs that can be generated in HMSC exemplified on coastal zoobenthos data from the Baltic Sea, but applicable and relevant to all community types

and habitats. The results focus on the inclusion of functional traits in community analysis and the species interaction network, to illustrate species associations. Species-to-species interactions, and hence their associations ultimately depend on the set of functional traits they express. Not being fully included yet, one future task will be to better incorporate traits into the species interaction network structure (Figure. 1b). However, there is already now a huge potential to predict geographic ranges of community-weighted traits in relation to their environment that could

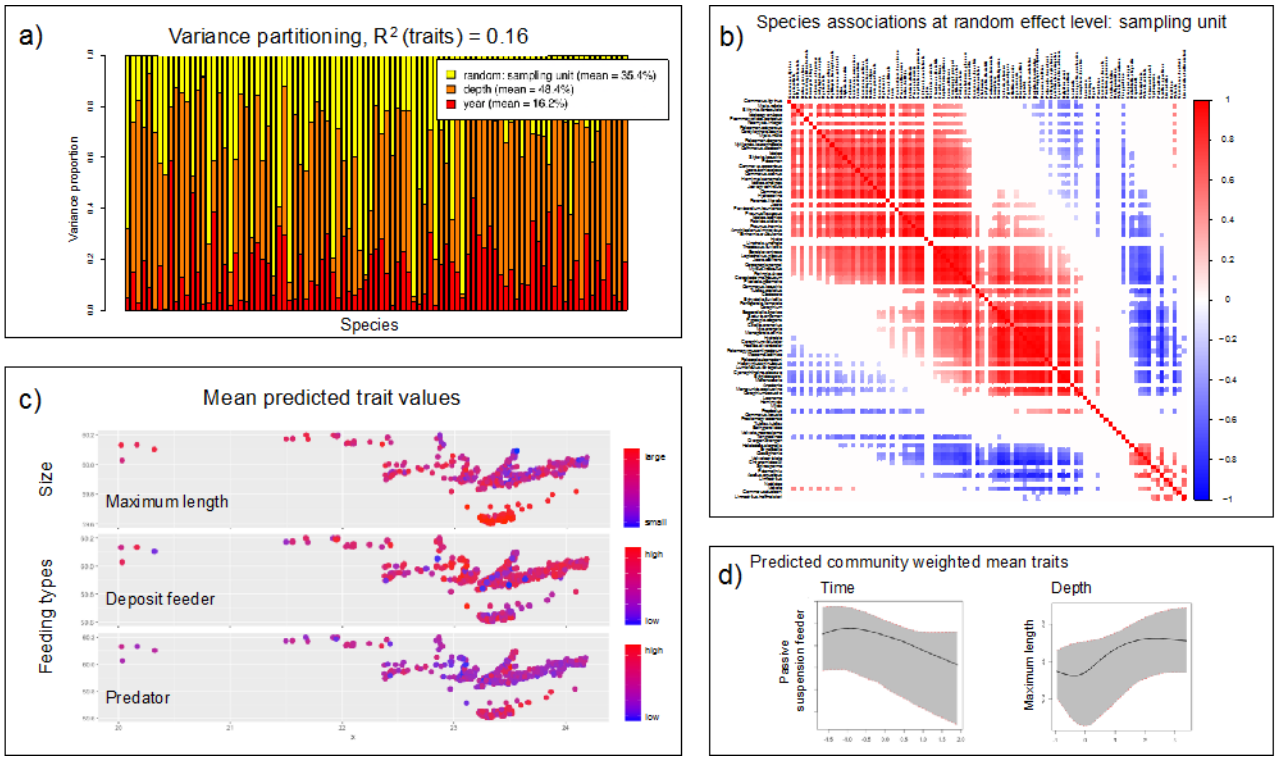


Figure 1. **a**) **Results of variance partitioning**. Variation in species occurrence is partitioned into responses to fixed and random covariates. Fixed effects include depth and year, with the sampling unit set as the random effect. The bar-plot shows the species-specific results whereas the legend indicates averages over all species. Traits explain 16 % of the fixed effect. **b) Estimates of species associations measured by residual correlation**. Species-to-species association matrices highlight species pairs, with positive associations indicated in red and negative ones in blue, here with statistical support of at least 75% posterior probability (the remaining cases are shown in white). Species are ordered to emphasize the network structure. **c) Mean predicted trait values of zoobenthos assemblages for each sampling site,** here exemplified for the traits: size, as maximum length and two feeding types, deposit feeders and predators. Blue indicates low and red indicates high values. **d) Predicted community weighted mean of traits as response to included covariates**, here exemplified with the predicted development of passive suspension feeders over time, showing a general decrease, and the maximum length against depth, which indicates that species tend to be bigger in deeper waters. Both predictions show high confidence intervals and can here only point out broad trends for a general illustration of the approach.

support conservation and management programmes. Especially being able to predict the impact of changing environmental factors on functional aspects of communities will be a key element to better understand the effect of climate-induced changes on ecosystem functioning and services. HMSC provides a framework for the inclusion of functional traits in community analysis as a hierarchical Bayesian joint species distribution model (Ovaskainen et al., 2017) and is implemented in R and a Matlab package.

References

Brey, T., Müller-Wiegmann, C., Zittier, Z.M.C., and Hagen, W. 2010. Body composition in aquatic organisms — A global data bank of relationships between mass, elemental composition and energy content. J. Sea Res. 64: 334-340.

Bruggeman, J., Heringa, J., and Brandt, B.W. 2009. PhyloPars: estimation of missing parameter values using phylogeny. Nucleic Acids Res. 37 (Web Server issue): W179-84. doi: 10.1093/nar/gkp370.

European Marine Observation Data Network (EMODnet) Biology project (www.emodnet-biology.eu), funded by the European Commission's Directorate - General for Maritime Affairs and Fisheries (DG MARE). Available online at www.emodnet-biology.eu

Finnish Environment Institute SYKE; 2018; Finnish Baltic Sea benthic monitoring, POHJE database

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. and Abrego, N. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett. 20: 561–576.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. and Hui, F.K.C. 2015. So Many Variables: Joint Modeling in Community Ecology. Trends Ecol. Evol. 30: 766–779.

Webb, T. et al. Body sizes of UK marine species, draft dataset for MERP v2017_05_09. Supplemented with data provided by Beth Mindel on 29/6/2017.

Webb, T., and Hosegood, J. 2013. Integrating biological traits of European marine benthic taxa into the World Register of Marine Species. EMODnet Project Final Report. http://www.marinespecies.org/traits/docs/reports_pilots/Benthos.pdf

Weigel, B., Blenckner, T. and Bonsdorff, E. 2016. Maintained functional diversity in benthic communities in spite of diverging functional identities. Oikos 125: 1421–1433.

3.6. GlobENV: Towards a High-resolution Climatology for the Seafloor

Andrew J. Davies[1,2] and Emyr M.T. Roberts[1,3]

[1]School of Ocean Sciences, Bangor University, Menai Bridge, Anglesey, Wales.

From August 2018: [2]Department of Biological Sciences, University of Rhode Island, Kingston, USA, davies@uri.edu.

[3]Department of Biological Sciences & KG Jebsen Centre for Deep-Sea Research, University of Bergen (UiB), Norway

Rationale

The ocean is the largest habitat on earth, covering approximately 70% of the planet. Our knowledge of patterns within surface waters is fairly extensive, principally driven by the development of earth observing satellites. Whilst marine scientists have clearly benefitted from such technologies, effectiveness is limited to the upper parts of the water column (e.g., ocean colour: Behrenfield and Falkowski, 1997) or coarse gravity estimates of the sea-surface that correlate with ocean depth (Smith and Sandwell, 1997). Accurate data regarding conditions at the seafloor remain scarce and are generally concentrated around developed countries (Ramirez-Llodra et al., 2010). Even with the now widespread adoption of technologies such as multibeam echosounders, remotely operated vehicles and autonomous underwater vehicles (Danovaro et al., 2014), only approximately 5% of the seafloor has been mapped, and a far smaller area has been investigated in great detail (Ramirez-Llodra et al., 2010). Recently, there has been renewed interest in ocean exploration, driven by the need to have a better understanding of geological features, underwater resources and species distributions. However, studies in many parts of our ocean remain constrained by the availability of high quality and validated data on seafloor conditions. Large-scale ocean mapping requires significant infrastructure and investment, and as a result, our understanding of the ocean floor significantly lags behind terrestrial environments.

Towards a new deep-sea climatology

Several marine climatologies are currently available. For example, Bio-ORACLE initially provided a data package that focusses on surface waters (Tyberghein et al., 2012) which was recently extended to include some benthic and future climate data (Assis et al., 2017). The MARSPEC dataset was based upon a higher resolution bathymetric dataset (i.e., SRTM30 Becker et al., 2009), and provides several benthic terrain variables and temperature/salinity for the sea surface (Sbrocco and Barber, 2013). GlobENV aims to extend these previous climatologies by providing an up-scaling approach that can be applied to any bathymetric dataset available of any resolution by using the best available environmental data in the ocean. It extends previously upscaled datasets (Davies and Guinotte, 2011; Guinotte and Davies, 2014), by providing a more

accurate and computationally efficient methodology and by implementing a robust validation framework that can be applied to each variable.

Methodology

Trilinear interpolation on a three-dimensional regular grid was used to estimate conditions on the seafloor from various environmental data sources. This approach interpolates the value of a point at depth *z* with coordinates *x* and *y* from eight surrounding points obtained from regularly gridded environmental data (Figure 1). GlobENV is designed to take gridded environmental data of varying resolution, and bathymetric data of an equal or finer resolution, producing an output that is an estimate of the variable at that particular depth and position. Finally, after computation of a variable (for an example see Figure 2), a validation process is conducted that compares performance against various environmental data. Several key parameters are calculated, 1) overall root-mean-square error, 2) spatial error calculations, 3) error by depth bin and 4) correlation metrics.

Conclusion

To date, GlobENV variables have been created for 10 different bathymetric datasets (5 global and 5 regional), and 7 environmental variables extracted from the World Ocean Database. Several more are planned, and both code and data products will be released under an open source license in the near future. Overall, the validation of the various variables indicates that finer resolution bathymetry produces a consistently lower RMSE than coarse bathymetric products, indicating that multibeam data would have particular value when used with GlobENV. However, the approach does suffer from several limitations as observed in Davies and Guinotte (2011). Depth and spatial error validation shows that the approach does not resolve well in shallower coastal seas, but performance improves rapidly at depths below 50 m irrespective of the bathymetric layer used. As such, this climatology may have particular value for the study of deep-sea species at varying resolutions depending on the data available.

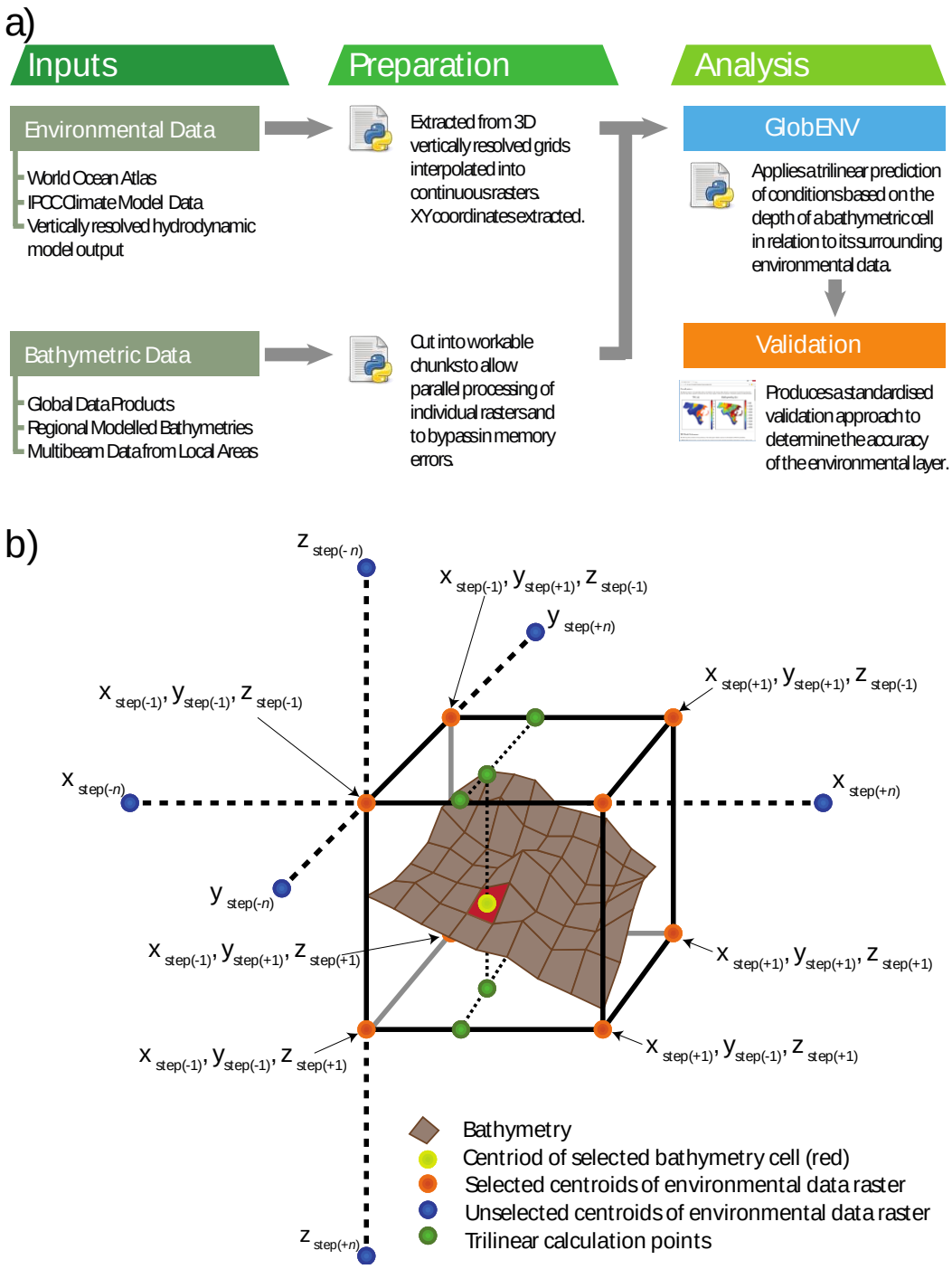


Figure 1. GlobENV workflow (a) and a schematic of the trilinear interpolation technique (b).

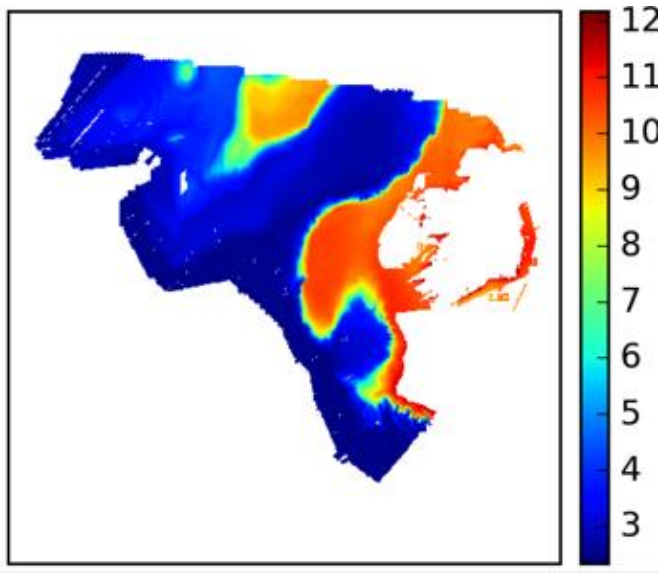


Figure 2. Example output from GlobENV at approximately 100 m XY grid resolution for the variable temperature (°C) based INFOMAR bathymetry and World Ocean Atlas 2013 data.

References

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrao, E.A. and De Clerck, O. 2017. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modeling. Global Ecol. Biogeogr. 27:277–284.

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G. and Weatherall, P. 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. Mar. Geod. 32: 355–371.

Behrenfield, M.J., and Falkowski, P.G. 1997. Photosynthetic reates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. 42: 1–20.

Danovaro, R., Snelgrove, P.V. and Tyler, P. 2014. Challenging the paradigms of deep-sea ecology. Trends Ecol. Evol. 29: 465–475.

Davies AJ, and Guinotte, J.M. 2011. Global habitat suitability for framework-forming cold-water corals. PloS One 6:e18483.

Guinotte, J.M., and Davies, A.J. 2014. Predicted Deep-Sea Coral Habitat Suitability for the U.S. West Coast. PLoS ONE 9(4): e93918. https://doi.org/10.1371/journal.pone.0093918.

Ramirez-Llodra, E., Brandt, A., Danovaro, R., De Mol, B., Escobar, E., German, C. R., Levin, L. A., Martinez Arbizu, P., Menot, L., Buhl-Mortensen, P., Narayanaswamy, B. E., Smith, C. R., Tittensor, D. P., Tyler, P. A., Vanreusel, A., and Vecchione, M. 2010. Deep, diverse and

definitely different: unique attributes of the world's largest ecosystem. Biogeosciences 7: 2851–2899.

Sbrocco, E. J., and Barber, P. H. 2013. MARSPEC: Ocean climate layers for marine spatial ecology. Ecology 94: 979.

Smith, W.H.F., and Sandwell, D.T. 1997. Global seafloor topography from satellite altimetery and ship depth soundings. Science 277: 1957–1962.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. 2012. Bio-ORACLE: A global environmental dataset for marine species distribution modeling. Ecol. Biogeogr. 21:272–281.

3.7. Ensembling of Multiple Data Sets and Multiple Models

Chris Rooper[1*], Duane Stevenson[1], Ivonne Ortiz[2], Jerry Hoff[1]

[1]Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA;

[2]Joint Institute for the Study of the Atmosphere and Oceans, University of Washington, Seattle, Washington, USA

*Current affiliation: Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

One of the advantages of using species distribution models is their ability to take advantage of diverse data sources using a variety of different methods. The data used for modeling the distribution of species in the deep-sea can range from statistically robust and well-designed surveys (e.g., Laman et al., 2018) to compilations of data from museum specimens, opportunistic collections or multiple sources (e.g., Davies and Guinotte, 2011). Modeling methods can incorporate different types of records, such as presence-only data (e.g., maximum entropy models) or presence/absence and abundance records using either statistical (e.g., generalized linear models and generalized additive models) or computer-learning (e.g., random forest and boosted regression tree models) methods. Species distribution models can also account for spatial autocorrelation explicitly in the method (e.g., kernel density models and vector auto-regressive spatial-temporal models), implicitly (e.g., maximum entropy models) or not at all. The choice of which data to use in the modeling and which modeling method to pursue can sometimes result in different representations of the species distribution (Figure 1).

Often ensembling data into a species distribution model can be problematic. The most common example of this in the deep-sea environment is combining data collected with different gear types

into a single modeling framework. This is routinely done with maximum entropy models, but can also be achieved with statistical and computer learning models where a term for gear type is included in the model formulation. An example of this can be found in vector-auto-regressive-spatial-temporal (VAST) models where a separate "catchability coefficient" can be set for each gear type (Thorson et al., 2015).

Another example of using ensembled data in species distribution models is for models predicting future changes in species distribution due to climate change effects. In this case there may be multiple climate scenarios each providing a time series of future temperatures. These data could be fed into a single model to produce multiple potential representations of species distributions. An example of this is shown for multiple climate scenarios acting on Pacific cod distributions in the eastern Bering Sea (Figure 2) or Araujo and New (2007).

Ensemble models have seen wide use in climatology, where multiple scenarios are combined to come up with the best predictions attainable. Ensembling is also becoming more prevalent in the fisheries stock assessment realm, where multiple types of assessment models with different assumptions can be combined into a trend in population status (e.g., Anderson et al., 2017 and Rosenberg et al., 2018 papers on "super-ensembles"). In the species distribution modeling literature there are a number of examples of using multiple models in an ensemble (Robert et al., 2016; Rowden et al., 2017; Rooper et al., 2017). The methods used to combine models into an ensemble have varied, but most are weighted by some type of measure of variability in the goodness-of-fit (e.g., AUC, SD, or R^2). In some cases, some sort of scaling of predictions may be necessary to combine model types into an ensemble. A particular subset of ensembling, that is interesting and relatively underutilized, is spatially explicit model ensembles. In some cases models have been constructed at different spatial scales, for example a PICES working group has been conducting broad-scale modeling of corals and sponges for the entire North Pacific ocean. This is an effort that is fairly data poor for some regions and uses presence only data. However, within the North Pacific Ocean there are a number of EEZ's and seamounts for which better models have been produced. For example, Miyamoto et al. (2017) constructed models for the Emperor Seamount chain. It would be beneficial to ensemble these smaller scale and larger scale models to come up with better overall predictions. Importantly, this type of ensemble should somehow represent spatially explicit confidence estimates for the predictions.

Key questions for ensembling data and models are:

- What are the advantages and disadvantages of combining data sets into a single modeling method versus modeling each data set independently and combining predictions?
- How to best weight or scale models when ensembling (especially across different data types)
- How to represent error in ensemble models that reflects individual model prediction variability

An example of ensembling methods with different data and models can be found with the distribution of skate nursery areas in the eastern Bering Sea. Several species of skates in the eastern Bering Sea deposit eggs in highly concentrated areas that are persistent through time (Hoff, 2010). The skate eggs incubate and develop for up to several years before a small but fully developed juvenile emerges. The persistence and concentration of eggs in specific nursery locations has been well documented, but the distribution of the nursery areas is relatively unknown. The data set for these nurseries is limited to 26 locations that have been discovered since 2008 during camera surveys and bottom trawl surveys. Recorded absences from the same camera and trawl surveys can be used, however, the catchability of skate eggs is presumed to be low, so these are probably not all "true" absences.

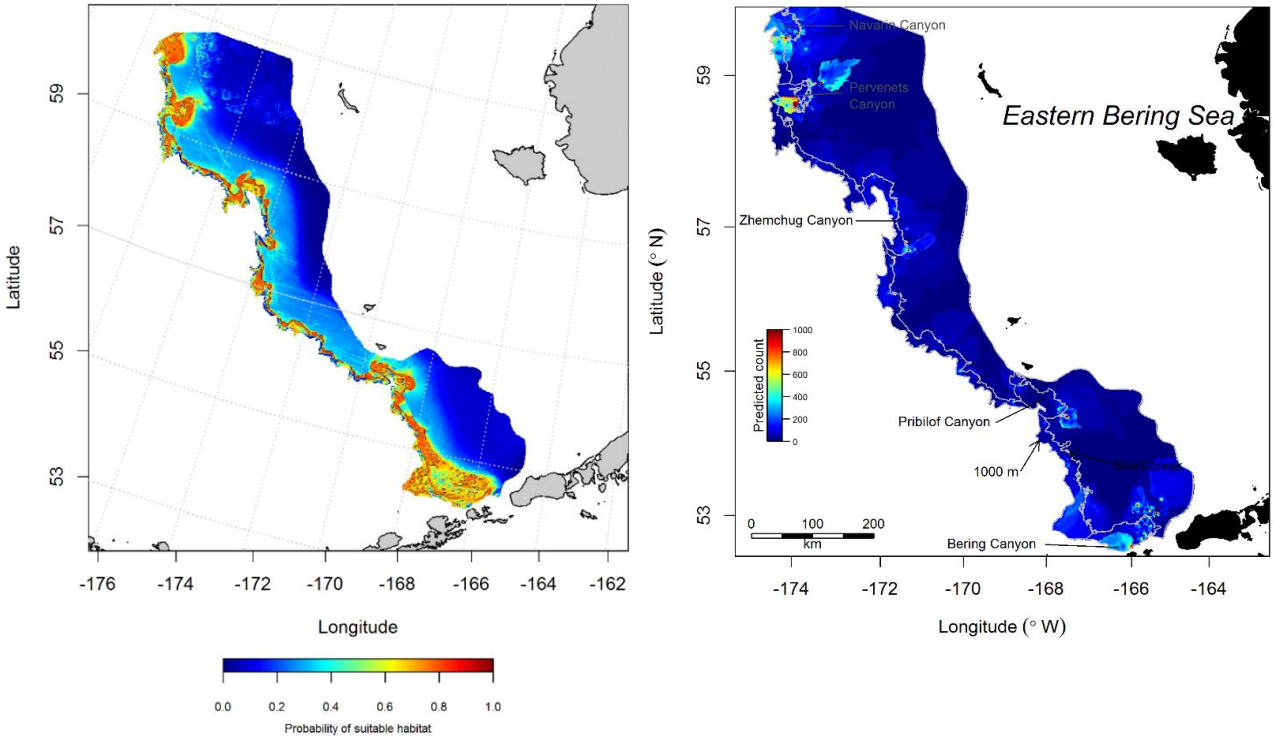


Figure 1. A) Map of the best model of probability of suitable habitat for skate nursery areas based on environmental variables, presence observations (n = 26), and absences for the eastern Bering Sea outer shelf and slope. The contour line indicates the threshold probability (0.60) at which presence of suitable habitat was determined. B) Kriged surface of the number of skate egg cases in the commercial fisheries catch across multiple gears, seasons and years. The data was collected by observers onboard the fishing vessels.

A second data set is available on the distribution of skate egg cases recorded by observers on commercial fishing vessels. This data set includes abundance (counts) and locations of catches of skate eggs. There are no absence records for these data and because of confidentiality issues and the relatively coarse position data available for the presence records, they are resolved on a scale

of ~5-10 km. These data are very different to the bottom trawl survey catches and camera survey data.

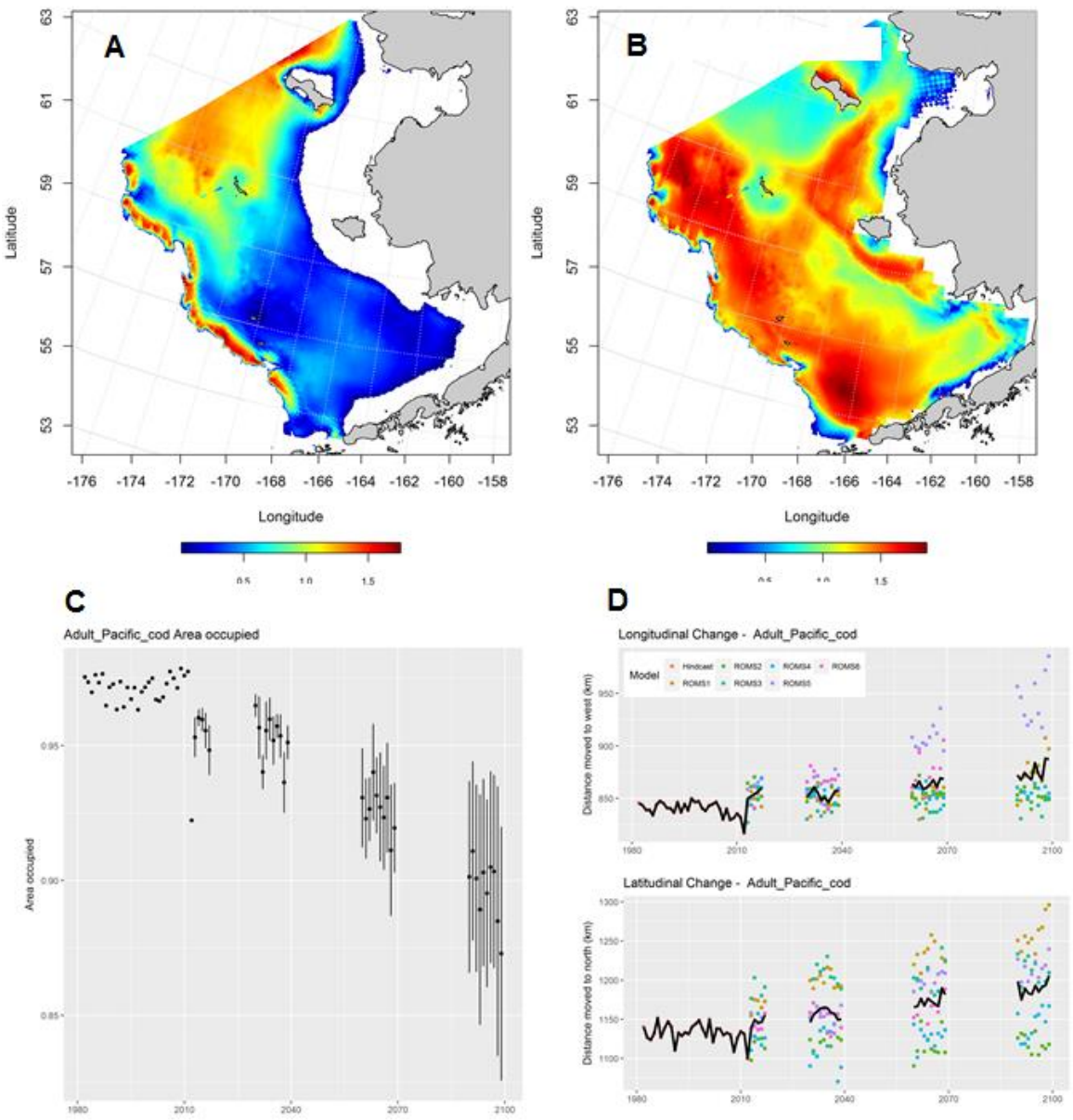


Figure 2. Predictions from a species distribution model for Pacific cod in the eastern Bering Sea using 6 climate forecasts. A) shows the distribution of Pacific cod in 2016, B) shows the predicted distribution of Pacific cod in 2099, C) shows the reduction (average and standard deviation) in area occupied by Pacific cod from 1980 to 2099 and D) shows the latitudinal and longitudinal shifts of the center of distribution for each of the climate scenarios.

In response to management concerns species distribution models were developed that predicted potential nursery locations based on the 26 known areas and the fisheries data. Combining the two

data sets was unreasonable given their fundamentally divergent properties, thus two modeling approaches were used. A maximum entropy model was used for the camera and bottom trawl surveys (with recorded absences as pseudo-absences). This model resulted in a fairly narrow depth band along the upper continental slope that was predicted to be suitable habitat (Figure 1). For the commercial fisheries data a kriging method was used that incorporated the abundance data.

These two outputs (Figure 1) reflect similar patterns in some cases, but are fundamentally different. How can these outputs best be combined in an ensemble to inform management on where skate nurseries are likely to be found?

References

Anderson, S.C., Cooper, A.B., Jensen, O.Pl, Minto, C., Thorson, J.T., Walsh, J.C., Afflerbach, J., Dickey-Collas, M., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Mosqueira, I, Rosenberg, A.A. and Selig, E.R. 2017. Improving estimates of population status and trend with superensemble models. Fish Fish. 18: 732-741. https://doi.org/10.1111/faf.12200

Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22: 2–47. doi: 10.1016/j.tree.2006.09.010

Davies, A.J., and Guinotte, J.M. 2011. Global habitat suitability for framework-forming cold-water corals. PLOS One 6:1-15.

Hoff, G.R. 2010. Identification of skate nursery habitat in the eastern Bering Sea. Mar. Ecol. Prog. Ser. 403: 243-254.

Laman, E.A., Rooper, C.N., Turner, K., Rooney, S., Cooper, D.W., and Zimmermann, M. 2018. Using species distribution models to define essential fish habitat in Alaska. Can. J. Fish. Aquat. Sci. Early online. https://doi.org/10.1139/cjfas-2017-0181

Miyamoto, M., Kiyota, M., Murase, H., Nakamura, T., and Hayashibara, T. 2017. Effects of bathymetric grid-cell sizes on habitat suitability analysis of cold-water gorgonian corals on seamounts. Mar. Geod. 40: 205-223. DOI: 10.1080/01490419.2017.1315543

Robert, K., Jones, D.O.B., Roberts, M.J., and Huvenne, V.A.J. 2016. Improving predictive mapping of deep-water habitats: considering multiple model outputs and ensemble techniques. Deep Sea Res. I 113: 80–89 doi: 10.1016/j.dsr.2016.04.008

Rooper, C.N., Zimmermann, M., and Prescott, M. 2017. Comparisons of methods for modeling coral and sponge distribution in the Gulf of Alaska. Deep Sea Res. II 126:148-161.

Rosenberg, A.A., Kleisner, K.M., Afflerbach, J., Anderson, S.C., Dickey-Collas, M., Cooper, A.B., Fogarty, M.J., Fulton, E.A., Gutierrez, N.L., Hyde, K.J.W., Jardim, E., Jensen, O.P., Kristiansen, T., Longo, C., MInte-Vera, C.V., Minto, C., Mosquiera, I., Osio, G.C., Ovando,

D., Selig, E.R., Thorson, J.T., Walsh, J.C. and Ye, Y. 2018. Applying a new ensemble approach to estimating stock status of marine fisheries around the world. Cons. Letters 11: 1-9, e12363.

Rowden, A.A., Anderson, O.F., Georgian, S.E., Bowden, D.A., Clark, M.R., Pallentin, A., and Miller, A. 2017. High-resolution habitat suitability models for the conservation and management of vulnerable marine ecosystems on the Louisville seamount chain, South Pacific Ocean. Front. Mar. Sci. 4: 1-19.

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. J. Cons. 72(5): 1297–1310.

3.7.1. Group Discussion

Dr. Rooper's presentation focused on the difficulties associated with ensembling data into species distribution modeling work and using this method to predict future changes in species distributions. Two comments were made concerning climate change predictions. First, it was suggested that it is important to understand the difference between climate predictions and climate projections. Projections are averaged statistical properties over a period of time, while projections deemed most likely become predictions or forecasts. Second, it was noted that one should use the same climate model when training data as when making predictions. It was suggested that it would be better to produce six models for six climate prediction data sets, and then ensemble the model output, rather than ensembling the data and producing a single output from that. Also, to obtain finer-scale climate predictions it was suggested that coarse climate change predictions could be upscaled by applying the data in "large pixels" to "finer pixels" which are available in present "nowcasts".

3.8. Resolution of Seabed Features in the Deep-Sea: Implications for Habitat Characterization

Myriam Lacharité and Craig J. Brown

Applied Oceans Research Group, Nova Scotia Community College, Dartmouth, NS, Canada

Seabed features – i.e., shape of the seafloor and substratum type - play a central role in habitat characterization for deep-sea benthos (e.g., Tong et al., 2012; Zeppili et al., 2016; and many others). Marine geomorphometry quantifies patterns of the shape of seafloor with derived bathymetric (i.e., 'terrain') variables (e.g., slope, roughness, curvature), which are also often used as proxy of substratum type, in particular in complex benthic ecosystems to predict the presence of hard substratum.

Using high-resolution seafloor data increases quality and accuracy of species distribution models (SDMs) in the deep-sea benthos (Rengstorf et al., 2012). Coarser environmental data can omit fine- (i.e., sub-grid) scale variability relevant to habitat characterization, introducing error into predictive models by mis-representing the distribution of a species' suitable habitat (Vierod et al., 2014). Because of data scarcity in the deep sea, SDMs are increasingly used at scales relevant to management (e.g., regional/territorial seas) to predict the occurrence of specific taxa of interest (e.g., cold-water corals) in unsurveyed regions. However, Vierod et al. (2014) suggest that the lack of accurate fine-scale bathymetric data limits the use of broad-scale models in the deep sea, and clear discrepancies between predicted and observed presence can arise during model validation (Anderson et al., 2016). While bathymetric gridded data are available at the global scale, their resolution remains coarse (30 arc-seconds: ~1 km at the equator; Weatherall et al., 2015), and the availability of higher-resolution bathymetric data from local surveys (sub-meter to 10s of meters) is sporadic (Mayer et al., 2018).

Here, we first suggest that in addition to fine-scale bathymetric data in the deep sea, care must be given to fine-scale substratum type, the characterization of which may not necessarily be captured using derived bathymetric variables. For example, recent studies have demonstrated the role of glacial dropstones (sporadic large boulders) at temperate and high latitudes in fostering diverse deep-sea benthos (Lacharité and Metaxas, 2017; Ziegler et al., 2017), the presence of which is captured with optical imagery or very-high bathymetric resolution (sub-meter). A detailed depiction of a wider extent of the surficial geology is detected with the reflectivity of the seafloor (i.e., acoustic backscatter), and advances in its standardization make these datasets increasingly more comparable (Brown et al., 2011; 2012).

Second, we suggest open and shared high-resolution bathymetric data available to multiple end-users would benefit habitat characterization for deep-sea benthos. The benefits of this approach for deep-sea SDMs goes beyond increased model accuracy by also supporting interdisciplinary research, in particular the development of high-resolution ocean hydrodynamic and biogeochemical models. Derived outputs from these models can enhance local habitat characterization and support spatially-explicit distribution models incorporating connectivity.

Finally, we support the view that broad-scale habitat characterization in the deep sea would benefit from the availability of multiresolution (spatially-varying) and multi-layered (bathymetry and backscatter) seabed data. Recent efforts aim to increase the bathymetric resolution at the global scale (Mayer et al., 2018). However, we suggest that high-resolution bathymetric datasets should be available wherever they have been collected, including the raw data and backscatter return for substratum characterization. Future research in deep-sea SDMs could explore how developing multiresolution modeling approaches would limit dependence of broad-scale high-resolution bathymetry.

References

Anderson, O.F., Guinotte, J.M., Rowden, A.A., Clark, M.R., Mormede, S., Davies, A.J., and Bowden, D.A. 2016. Field validation of habitat suitability models for vulnerable marine ecosystems in the South Pacific Ocean: Implications for the use of broad-scale models in fisheries management. Ocean Coast. Manag. 120: 110–126.

Brown, C.J., Sameoto, J.A., and Smith, S.J. 2012. Multiple methods, maps, and management applications: Purpose made seafloor maps in support of ocean management. J. Sea Res. 72:1-13.

Brown, C.J., Smith, S.J., Lawton, P., and Anderson, J.T. 2011. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar. Coast. Shelf Sci. 92: 502–520.

Lacharité, M., and Metaxas, A. 2017. Hard substrate in the deep ocean: How sediment features influence epibenthic megafauna on the eastern Canadian margin. Deep. Res. Part I Oceanogr. Res. Pap. 126: 50–61.

Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., Lamarche, G., Snaith, H., and Weatherall, P. 2018. The Nippon Foundation—GEBCO Seabed 2030 Project: The Quest to See the World's Oceans Completely Mapped by 2030. Geosciences 8: 63.

Rengstorf, A.M., Grehan, A., Yesson, C., and Brown, C. 2012. Towards high-resolution habitat suitability modeling of vulnerable marine ecosystems in the deep-sea: Resolving terrain attribute dependencies. Mar. Geod. 35: 343–361.

Tong, R., Purser, A., Unnithan, V., and Guinan, J. 2012. Multivariate statistical analysis of distribution of deep-water gorgonian corals in relation to seabed topography on the Norwegian margin. PLoS One 7: e43534.

Vierod, A.D.T., Guinotte, J.M., and Davies, A.J. 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep. Res. Part II Top. Stud. Oceanogr. 99: 6–18.

Weatherall, P., Marks, K.M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J.E., Rovere, M., Chayes, D., Ferrini, V., and Wigley, R. 2015. A new digital bathymetric model of the world's oceans. Earth Space Sci. 2: 331-345.

Zeppilli, D., Pusceddu, A., Trincardi, F., and Danovaro, R. 2016. Seafloor heterogeneity influences the biodiversity-ecosystem functioning relationships in the deep sea. Sci. Rep. 6: 1–12.

Ziegler, A.F., Smith, C.R., Edwards, K.F., and Vernet, M. 2017. Glacial dropstones: Islands enhancing seafloor species richness of benthic megafauna in West Antarctic Peninsula fjords. Mar. Ecol. Prog. Ser. 583: 1–14.

3.8.1. Group Discussion

Discussion following the presentation focused on the lack of high-quality broad-scale seabed data and possible methods of adjusting to this lack of data while still providing well-informed model outputs. Variables vary at different spatial scales and different variables are more or less useful as predictors depending 1) on the species/habitat whose distribution is being modeled, and 2) the questions that are being asked of the model (i.e., the model's intended use). Fine-scale data can be highly relevant to a given species' distribution, yet data at such a resolution may not (and likely won't) be available. One solution to this problem could be the use of proxy data, for example using satellite-based glacier tracking as a proxy for dropstones. A second method mentioned creating a substrate predictive map for small areas such as the coast of British Columbia; using substrate data from small areas to predict other areas is a good option when fine-scale data is unavailable.

Finally, it is important to distinguish between the scale at which the habitat varies compared to the scale at which the biota is sampled. Multiscale models were suggested as a possible answer for this problem. An additional example was given: it would be expected that rock would vary at a fine scale compared to other covariates that vary at large scales. If available data is limited, such as specific rock location, certain habitats can be eliminated for certain species. Models are conditional on covariates, and if these are absent it will show in the error.

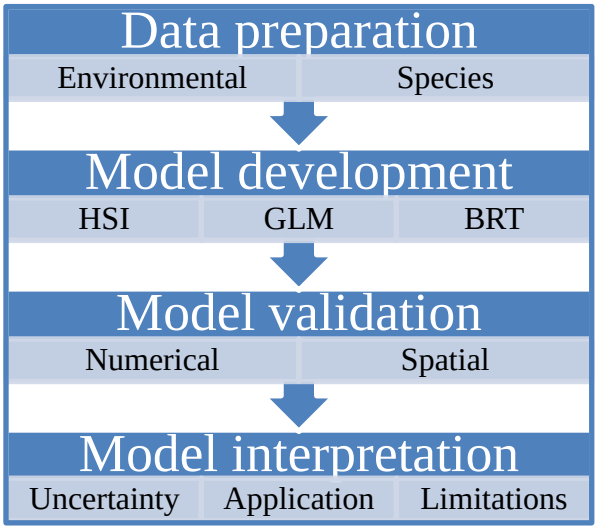
3.9. Best Practices in the Development and Application of Species Distribution Models to Support Decision Making in Marine Spatial Planning

Jessica Finney[1], Emily Rubidge[1], Cherisse Du Preez[1], Jessica Nephin[1], Candice St. Germain[1], Cole Fields[1], and Edward Gregr[2]

[1] Pacific Region, Fisheries and Oceans Canada (DFO)
[2] SciTech Consulting, Vancouver, BC

The Marine Spatial Ecology and Analysis (MSEA) Section is a part of Fisheries and Oceans Canada (DFO) in the Pacific Region. The MSEA Section is involved in several areas of research that use SDMs to support ocean management, including marine protected area network design, oil spill response planning, forecasting the impacts of climate change, and identifying sensitive benthic areas. The Species Distribution Modeling Working Group (SDM WG) was formed to provide coordination and advice to help facilitate the use and development of species distribution models within MSEA and more generally in DFO in the Pacific Region.

A key output of the SDM WG will be a **best practices workbook**. The intent of this workbook is to provide users with a well-referenced guide to the development of species distribution models. The workbook will include discussions on appropriate data usage, pre-processing and data preparation, modeling approaches and development, model validation, and interpretation of results. The workbook will outline best practices to standardize the species distribution modeling approach within the MSEA section (to ensure consistent quality and rigor in the models developed in the section). It will also include important information about how to interpret model outputs and how to convey underlying uncertainty to managers. The workbook will facilitate and simplify the model building and interpretation process, support the use of best practices, and provide consistent outputs, measures of uncertainty, and evaluation metrics for managers.



The workbook will present three types of commonly used models of increasing complexity and data needs to highlight the importance of the four basic steps of species distribution models development (data preparation, model development, model validation, and interpretation) and to identify when a simple approach may be more appropriate than a complex approach. The three models that will be included in the workbook to demonstrate these points are: 1) **Habitat Suitability Index (HSI)** models (simple model); 2) **Generalized Linear Models (GLMs)** (medium complexity); and 3) **Boosted Regression Tree (BRT)** models (most complex). This tiered approach increases the utility of the workbook by making it applicable to a wide range of data availability scenarios. It also lets users decide how complex they need or want their models to be. For example, if a simple model provides good results, it may not be necessary to spend the time and effort required to build a more complex model.

An additional component of the workbook will be automated code for both the GLM and BRT models. The species distribution modeling procedure for both models is run in two stages. The first stage processes species (presence/absence or abundance) and environmental data, produces diagnostic plots, and maps environmental data. The second stage builds and evaluates the GLM and BRT models using k-fold cross validation, produces validation plots, estimates the marginal effects and relative influence of predictor variables in the model and uses models to predict the presence or abundance of species in new locations.

The workbook will initially be tested out using a set of 12 benthic marine species representing a variety of life-history strategies, habitats and data availability scenarios. Results of this analysis, as well as the completed workbook, are expected in early 2019.

3.10. Evaluating Effects of Rescaling and Weighting Data on Habitat Suitability Modeling

Ying Xue[1,2], Lisha Guan[2], Kisei R. Tanaka[2], Zengguang Li[1,2], Yong Chen[2], Yiping Ren[1]

[1] Fisheries College, Ocean University of China, Qingdao 266003, China
[2] School of Marine Sciences, University of Maine, Orono, ME 04469, USA

Background

Habitat suitability index (HSI) models are important tools in identifying suitable habitats of living marine resources (LMR). Abundance indices (AI) derived from fishery-independent surveys are often used to calibrate HSI models. However, AIs tend to have a highly right-skewed distribution as a result of large spatial heterogeneity of LMR distributions, which can further affect subsequent HSI model performance. Furthermore, traditional HSI models are often based on unrealistic assumptions that environmental variables have equal impacts on the AIs. Using American lobster (*Homarus americanus*) in the inshore Gulf of Maine as a case species, this study evaluates the performance of different approaches in developing HSI models.

Objectives

The objectives of this study are to compare the performance of HSI models derived using untransformed abundance indices (AIs) versus log-transformed AIs and evaluate the effectiveness of a boosted regression tree (BRT)-based weighting approach.

Approaches

The existence of large AI values might result in underestimation of HSI values for most sampling stations. One approach to avoid this problem is to use rescaled AIs (e.g., log-transformed AIs) to reduce the impact of large AI values on SIs. In this study, boosted regression tree (BRT) models were used to determine the weights of environmental variables in HSI modeling.

Key findings

Both cross-validation and predicted habitat suitability maps suggested that the weighted HSI model based on log-scaled AI data tended to yield a more reliable prediction of optimal habitats for American lobster (Figure 1). The unweighted HSI model based on the original AI data, however, tended to underestimate optimal habitats and overestimate suboptimal habitats. We recommend using log-transformed AIs and determining the weights of different environmental variables based on the BRT method in HSI modeling, especially when AI data are highly skewed.

The approach demonstrated in this study can improve the quality of HSI modeling, leading to better definitions of suitable habitats (Xue et al., 2017).

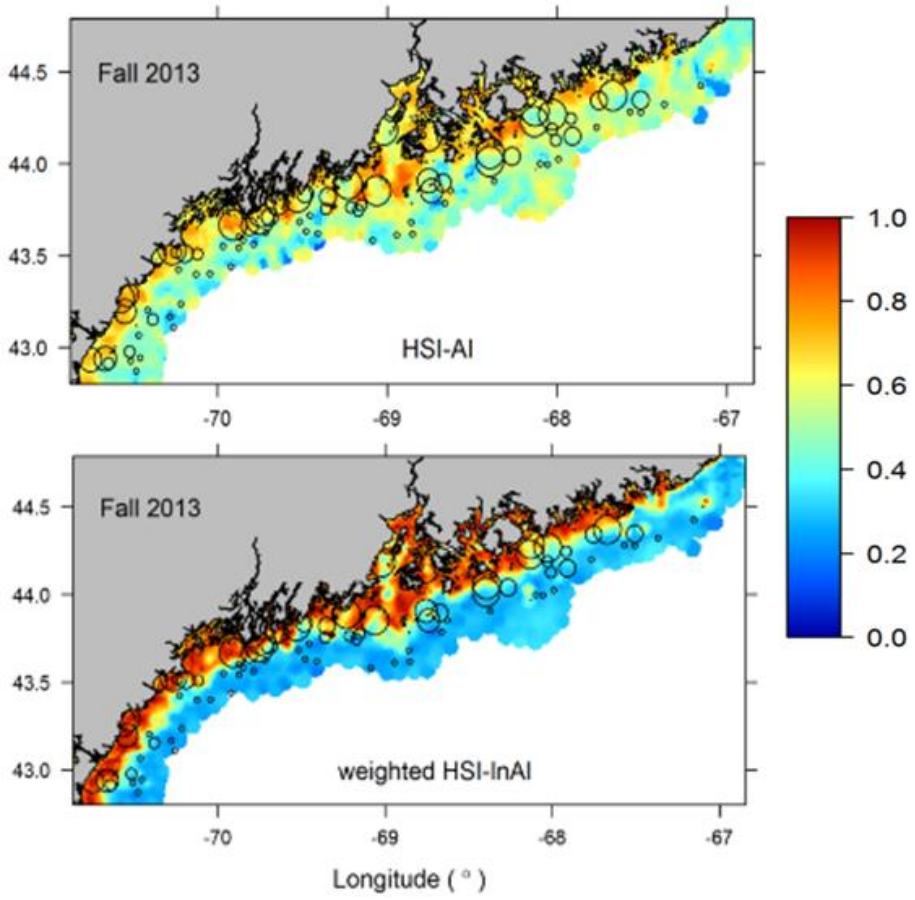


Figure 1. Spatial distribution of observed abundance indices (AIs) for American lobster in fall of 2013 overlaid on two predicted habitat suitability index (HSI) maps derived from HSI-AI and weighted HSI-lnAI models in Maine–New Hampshire inshore bottom trawl survey areas.

Reference

Xue, Y., Guan, L., Tanaka, K., Li, Z., Chen, Y., and Ren, Y. 2017. Evaluating effects of rescaling and weighting data on habitat suitability modeling. Fish. Res., 188: 84–94. https://doi.org/10.1016/j.fishres.2016.12.001

3.10.1. Group Discussion

The discussion following Dr. Tanaka's presentation focused on the use of a simple rank-based (i.e., non-statistical) Habitat Suitability Index (HSI) model to predict species

distribution. It was pointed out that this type of qualitative method can be useful for introducing non-statisticians to the field of species modeling – this is particularly important in helping non-modeller scientists who engage with stakeholders and need to understand species distribution models to a sufficient degree to simply explain the methodologies and their outputs to stakeholders. One problem with this method, however, is that when calibrating HSI with highly skewed zero-inflated data, species abundance and environmental variables may not be associated appropriately in terms of computing the habitat suitability of marine species. Furthermore, if environmental variables are weighted equally, the model output can be problematic as it ignores the difference in ecological importance of each environmental variable. The presentation showed that rescaling skewed observations (e.g., log-transformation) and applying machine-learning techniques (e.g., boosted regression tree) can address these known shortcomings in the HSI modeling. These approaches are important because weighted environmental variable contributions are generally an output which is sought after from species distribution modeling work, rather than used as an input.

3.11. Developing a Generalized Climate-niche Modeling Framework to Improve Management of Commercially Important Species in a Climatically Altered Marine Environment: A Case Study with American lobster and Atlantic scallop in the Gulf of Maine

K. R. Tanaka*[1,2], M. P. Torre*[1], V. S. Saba[3], and Y. Chen[1]

*Both authors contributed equally

[1]School of Marine Sciences, University of Maine, Orono, ME 04469, USA
[2]Atmospheric and Oceanic Sciences Program, Princeton University, 300 Forrestal Road, Sayre Hall, Princeton, New Jersey 08544, USA
[3]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, c/o Geophysical Fluid Dynamics Laboratory, 201 Forrestal Road, Princeton University Forrestal Campus, Princeton, New Jersey 08540, USA

Background

Climate change has been identified as a key driver of ecology and population dynamics of commercially important fish and shellfish stocks within marine ecosystems. Accounting for changing biogeography of fish populations is essential to the effective assessment and management of fisheries; however, the majority of fishery management frameworks in the United States have yet to incorporate information about how environmental conditions influence

distribution and abundance of commercial stocks (Saba et al., 2015; Skern-Mauritzen et al., 2015). That observed biogeographic ranges shift as consequences of ongoing climate change has raised a lot of concerns and uncertainty over the effectiveness of the current survey programs and subsequently in associated stock assessment models. The development of a forecasting ability to evaluate climate change impacts on the ecology of commercially valuable marine stocks has been advocated as an important step in mitigating uncertainty in climate change adaptation strategies.

Objectives

Using two commercially important stocks (American lobster and Atlantic scallop) in the northeast USA continental shelf as case studies, the objective of this research is to develop a generalized climate-niche modeling framework that can be a part of a coordinated regional effort to assess climate change effects on commercial fish and their fisheries. The proposed climate-niche modeling framework will improve our management of marine fish and shellfish stocks by providing (1) enhanced hind/now/forecasting capacity of spatio-temporal changes in their biogeography, and (2) a space/time varying estimate of their availability to the existing fishery surveys and management zones. Such a modeling framework will provide excellent tools for climate change-related research in fisheries, thus greatly enhancing foundations for future climate-change-related research proposals in fisheries.

Approach

The proposed climate-niche modeling framework will consist of using a delta-generalized linear mixed model (delta-GLMM: Thorson et al., 2015) and ensemble niche model (biomod2: Thuiller et al., 2016) coupled with a dispersion simulation (MigClim: Engler et al., 2012) for hindcasting/forecasting spatio-temporal changes in distribution of American lobster and Atlantic scallop on the northeast USA continental shelf. In this project we would like to address the following questions: (1) How did abundance and distribution of American lobster and Atlantic scallop change over time? (2) What are the management implications of changing biogeography of the lobster and scallop stocks to the existing fishery survey designs and management zones?

Preliminary results

The ensemble niche model was used to predict the change in spatial distribution of American lobster and Atlantic scallop based on an 80-year time series of future bottom temperatures and salinity changes based on a transient climate response ($2xCO^2$) simulation where atmospheric CO^2 is increased by 1% per year (Figures 1, 2). The model results will be used to (1) evaluate the past, present and future performance of the existing fishery-independent survey design in capturing annual variability of lobster and scallop stock abundance, and (2) discuss management implications of changing biogeography of the lobster and scallop stocks to the existing fishery management zones.

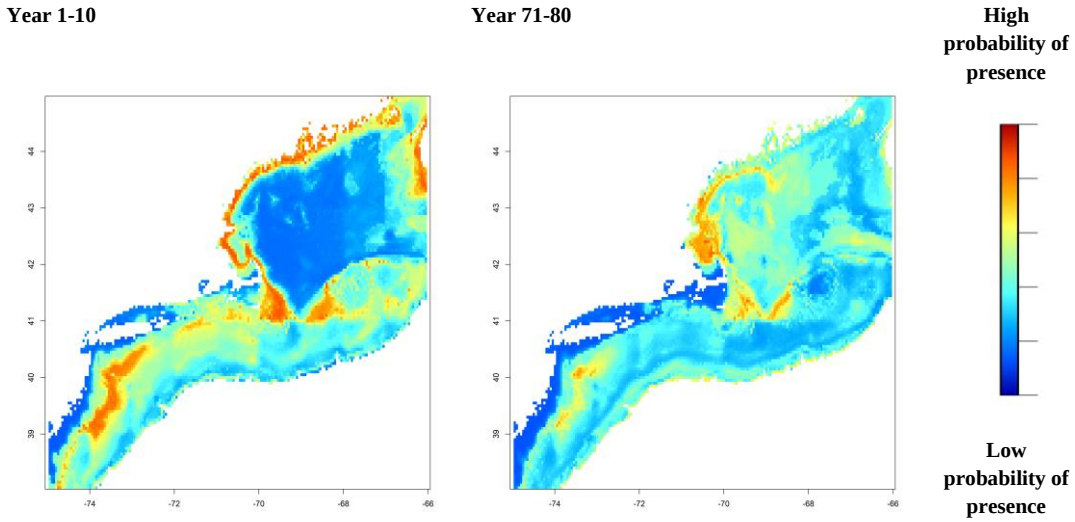


Figure 1. Predicting the change in American lobster probability of occupancy in the U.S.A. northeast continental shelf region. Predictions represent the mathematical average of ensemble models of probability of presence from years 1984–2015. Model predictions from the first and last 10 years are compared.

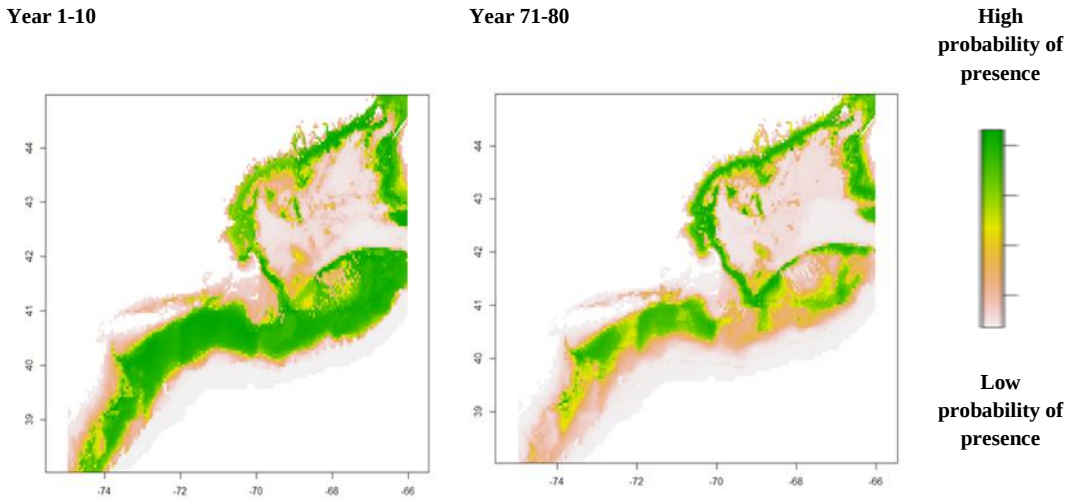


Figure 2. Predicting the change in sea scallop probability of occupancy in the U.S.A. northeast continental shelf region. Predictions represent the mathematical average of ensemble models of probability of presence from the years 1984–2015. Model predictions from the first and last 10 years are compared.

References

Engler, R., Hordijk, W., and Guisan, A. 2012. The MIGCLIM R package – seamless intergration of dispersal constraints into projections of species distribution models. Ecography 35: 872–878.

Saba, V.S., Griffies, S.M., Anderson, W.G., Winton, M., Alexander, M.A., Delworth, T.L., Hare, J.A., Harrison, M.J., Rosati, A., and Vecchi, G.A. 2015. Enhanced warming of the Northwest Atlantic Ocean under climate change. J. Geophys. Res. Ocean., 120: 118–132. https://doi.org/10.1002/2015JC011346.Received

Skern-Mauritzen, M., Ottersen, G., Handegard, N.O., Huse, G., Dingsør, G.E., Stenseth, N.C., and Kjesbu, O.S. 2015. Ecosystem processes are rarely included in tactical fisheries management. Fish Fish. 17: 165–175. https://doi.org/10.1111/faf.12111

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72: 1297-1310.

Thuiller, W., Georges, D., Engler, R., and Breiner, F. 2016. biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7. https://CRAN.R-project.org/package=biomod2, (2), 1–104. Retrieved from https://cran.r-project.org/web/packages/biomod2/index.html

3.11.1. Group Discussion

Following Dr. Tanaka's presentation, a suggestion was made that time-frames for predictions should be quantified scientifically rather than arbitrarily. For example, using timelines relevant to lobster fisheries stock assessments would be beneficial rather than picking 80-100 years to look at, simply because that timeframe is commonly used. Variables will likely change drastically on a smaller timeframe than 100 years, for example SST will likely change drastically on the 10-15 year timeframe. A second comment focused on scale: when a regional scale is used, more attention must be paid to individual variable responses. When a global scale is used, larger patterns must be looked at such as global fish decline. Finally, a third comment focused on how to quantify change. The Gulf of Maine was provided as an example, and how the baseline should be set was questioned, for example which date to start the analysis at when projecting into the next five or ten years.

3.12. Determining Thresholds for Interpretation of Probability Maps

Chris Rooper

Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA

*Current affiliation: Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

One of the primary uses of species distribution models is to predict where habitat for a deep-sea organism occurs and where it does not. This is a binary solution requiring a "yes" or "no" for each location of interest. Most species distribution models use data to predict a probability of presence or absence. The movement from probability of presence to presence or absence requires implementation of a threshold probability. Secondarily, many of the metrics used to evaluate performance of models are based on predictions of presence or absence (e.g., Cohen's Kappa, TSS).

The choice of a threshold for use in a management strategy can have impacts on the results that are communicated to decision makers. For example, Figure 1 shows an analysis of thresholds for four models related to eastern Bering Sea trawl fisheries. In this case, the question was asked "should we be protecting Pribilof Canyon from bottom trawling to protect deep-sea corals". An important decision point was how much of the total coral habitat was contained within the canyon. Three data sources were separately modeled (using two disparate methods) and an ensemble model was constructed as well. Using a number of different thresholding methods found in the literature, the proportion of coral habitat in Pribilof Canyon was predicted to range from ~17% to 90%. This example shows some of the major questions and concerns that arise when using thresholds with species distribution models. An incomplete list of these concerns that could be discussed are:

- Should we be doing this and are there alternatives?
- Which method should be chosen or how should the determination of method be made?
- Are there general guidelines that can be followed?
- How to best represent uncertainty in conclusions resulting from the choice of a specific threshold?

A growing body of literature has sought to evaluate some of these questions (e.g., Manel et al., 2001; Liu et al., 2005; Bean et al., 2011; Lawson et al., 2014). The consensus of most of these evaluations is that the best methods for thresholding are objective; meaning they set the threshold using statistics derived from the data (such as the prevalence approach) or the model predictions (such as sensitivity = specificity approach). However, it appears that the distribution of the modeled data and the modeling method are important considerations when choosing a thresholding method.

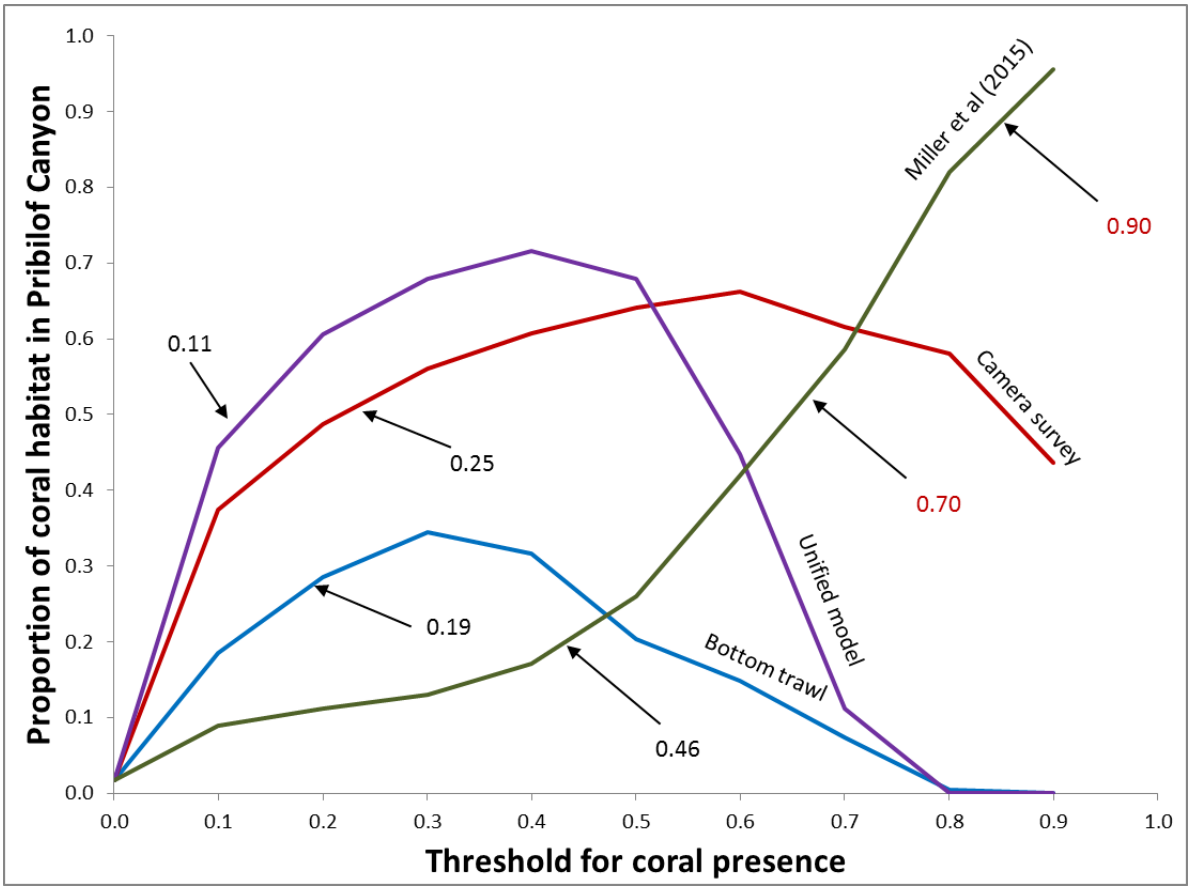


Figure 1. The proportion the total area of coral habitat in eastern Bering Sea that occurs in Pribilof Canyon predicted by four models as a function of the threshold probability for presence. Lines indicate the results for Generalized Additive Models of coral presence based on bottom trawl survey data (bottom trawl), camera survey data (camera survey) and both (Unified model) from Rooper et al. (2016) and a maximum entropy model based on camera data (Miller et al., 2015). The black arrows indicate the threshold where the predicted error rates for presence and absence are equal. The red numbers indicate two suggested thresholds based on high probabilities.

References

Bean, W.T., Stafford, R., and Brashares, J.S. 2011. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecogr. 35: 250–258

Lawson, C.R., Hodgson, J.A., Wilson, R.J., and Richards, S.A. 2014. Prevalence, thresholds and the performance of presence-absence models. Methods Ecol. Evol. 5: 54-64.

Liu, C.R., Berry, P.M., Dawson, T.P., and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecogr. 28: 385–393.

Manel, S., Williams, H.C., and Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38: 921-931.

Miller, R.J., Juska, C., and Hocevar, J. 2015. Submarine canyons as coral and sponge habitat on the eastern Bering Sea slope. Glob. Ecol. Conserv. 4: 85−94.

Rooper, C.N., Sigler, M., Goddard, .P, Malecha, P.W., Towler, R., Williams, K., Wilborn, R., and Zimmermann, M. 2016. Validation and improvement of species distribution models for structure forming invertebrates in the eastern Bering Sea with an independent survey. Mar. Ecol. Prog. Ser. 551: 117-130.

3.12.1. Group Discussion

The discussion following the presentation focused on thresholds, as well as model validation and uncertainty measures. To begin with, the concept of deriving a threshold without human selection was brought up, and whether it is better to use a mathematical based approach rather than a human-derived approach for thresholds. As well it was noted that the research question supporting the work will impact how conservative a threshold selection must be, whether it is more dangerous to have false positives or false negatives is an important consideration. When a threshold is selected, information is immediately lost, and this is important for managers and stakeholders to be aware of. An alternative approach was offered which suggested picking for example 80% of density (where 80% of observations will be occurring) rather than picking a threshold. Following this discussion, the concept of model validation was briefly covered. In order to calibrate a model, it must be validated, which can be expensive and unlikely for deep-sea situations. Independent data can be used for validation: often with a statistical model, available data is used to teach the model and then in order to validate it, cross-validation can be used rather than retrieving additional field data for validation. Finally, the concept of measuring uncertainty was covered. One question was raised concerning using data from one area to predict to another, and if these areas should be environmentally similar in order to do this. As well it was mentioned that a good idea is assigning a level of certainty to outputs based on which areas of the final prediction had input data compared to which areas were completely extrapolated towards, because uncertainty in the data will be translated to uncertainty in the model. In regards to the research question once more, which is better, 50% probability with 90% certainty, or 80% probability with lower certainty. This will depend upon the manager/stakeholder and what the model is being created for.

3.13. Some Uncertainties of Bathymetry Data

Chris Yesson

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY United Kingdom

Bathymetric data are important constituents of many habitat models, both directly in the form of seabed depth measurements and indirectly as constituents of other environmental layers. For example, oceanographic models have bathymetric grids as inputs, which help shape predictions of currents and temperature. Furthermore, we can directly infer seabed characteristics from bathymetry data by examining local variation in depth to create layers such as slope, aspect and terrain roughness, which are useful predictors of benthic organisms (e.g., Dolan et al., 2008). We can also examine the shape of the seabed to predict topographic features such as seamounts (e.g., Yesson et al., 2011), which are regarded as important habitats for many species (e.g., Bouchet et al., 2014). These topographic layers are often the highest resolution datasets upon which to base predictions of habitat suitability, and can be very important constituents of the models. However, it is important to consider uncertainties in the underlying bathymetry and consider the implications of this uncertainty on any resulting model.

Global bathymetry grids such as GEBCO (Weatherall et al., 2015) and SRTM (Becker et al., 2009) are widely used in large-scale habitat modeling studies. However, the grids themselves are models based on a combination of soundings (i.e., high resolution acoustic surveys) and satellite altimetry (lower resolution data from satellite sensors). Satellite- altimetry provides global coverage and is the foundation of bathymetry models, but these sensors cannot determine small features (i.e., seamounts under 1.5km, Wessel et al., 2010). Acoustic surveys generate data best suited for determining seabed depth and these are used to constrain models used to create bathymetry grids (Becker et al., 2009). Soundings are limited to a small proportion of the ocean and the majority of bathymetry grid data is derived from the underlying model, rather than acoustically surveyed. For example only 18% of GEBCO grid cells are directly supported by acoustic surveys (Weatherall et al., 2015). With so little sounding data available, there is a premium on making full use of the data available, and historical soundings (based on weighted lines) have been extracted from nautical charts to expand these data (Becker et al., 2009).

An example of the shortcomings of these data are presented as an anecdote. A survey was conducted in February 2016 in the British Indian Ocean Territory to visit seamounts around the Chagos Archipelago. Two shallow seamounts (summit <500m) predicted by Yesson et al., (2011), derived from a global bathymetric grid, were visited, but no seabed feature was detected above 1,500m. An examination of Admiralty charts for the area showed that the summits of these features had soundings reporting “no bottom detected at depth X” where the bathymetry grid (and subsequent seamount predictions) showed seamount summits at depth X. We believe these “no seabed” soundings have been misinterpreted as seabed depths in bathymetry grids such as GEBCO. In this case, spatially isolated, erroneously interpreted sounding data can lead to the false prediction of seabed features such as seamounts. We estimate that 12.5% of seamounts predicted to be in the British Indian Ocean Territory by Yesson et al., (2011) are “phantoms” based on falsely interpreted soundings.

Bathymetry grids are continually improving, whether that be from new multibeam acquisition, such as that collected during the search for Malaysian Airlines flight MH370 (Smith and Marks, 2014), or improved satellite gravity data (Sandwell et al., 2014). However, these bathymetry grids still rely on sparse sounding data for many regions. It is important that we consider the uncertainties in the bathymetry when using these to make predictive habitat maps.

References

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G. and Weatherall, P. 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. Mar. Geod. 32: 355–371.

Bouchet, P.J., Meeuwig, J.J., Salgado Kent, C.P., Letessier, T.B., and Jenner, C.K. 2014. Topographic Determinants of Mobile Vertebrate Predator Hotspots: Current Knowledge and Future Directions. Biol. Rev. 90: 699–728. https://doi.org/10.1111/brv.12130.

Dolan, M.F., Grehan, A.J., Guinan, J.C., and Brown, C. 2008. Modeling the local distribution of cold-water corals in relation to bathymetric variables: Adding spatial context to deep-sea video data. Deep Sea Res. Part I: Oceanogr. Res. Pap. 55: 1564-1579.

Sandwell, D.T., Muller, R.D., Smith, W.H.F., Garcia, E., and Francis, R. 2014. New Global Marine Gravity Model from CryoSat-2 and Jason-1 Reveals Buried Tectonic Structure. Science 346: 65–67. https://doi.org/10.1126/science.1258213.

Smith, W.H.F., and Marks, K.M. 2014. Seafloor in the Malaysia Airlines Flight MH370 Search Area. Eos Trans. AGU 95: 173–74. https://doi.org/10.1002/2014eo210001.

Weatherall, P., Marks, K.M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J.E., Rovere, M., Chayes, D., Ferrini, V., and Wigley, R. 2015. A New Digital Bathymetric Model of the Worlds Oceans. Earth Space Sci. 2: 331–45. https://doi.org/10.1002/2015ea000107.

Wessel, P., Sandwell, D., and Kim, S.S. 2010. The Global Seamount Census. Oceanogr. 23: 24–33. https://doi.org/10.5670/oceanog.2010.60.

Yesson, C., Clark, M.R., Taylor, M.L., and Rogers, A.D. 2011. The Global Distribution of Seamounts Based on 30 Arc Seconds Bathymetry Data. Deep Sea Res. Part I: Oceanogr. Res. Papers 58: 442–53. https://doi.org/10.1016/j.dsr.2011.02.004.

3.14. Mapping Uncertainty of SDM Predictions

Fiona Davidson and Anders Knudby

Department of Geography, Environment and Geomatics, University of Ottawa

Introduction

To help decision-makers use species distribution modeling for conservation planning, the accuracy of predictions made by the model must be estimated. While overall model accuracy is commonly evaluated using metrics such as the AUC, the uncertainty of predictions is better estimated spatially, especially when models are used to extrapolate predictions to a new area. The relative confidence in predictions across separate regions of the modeling extent can be assessed with a bootstrap method (Anderson et al., 2016; Rowden et al., 2017). This working paper presents the initial results of such a bootstrap method used to map uncertainty for a spatially extrapolated model.

Data

The data used for this project included hexactinellid sponge presence and absence data from the Pacific Ocean along the North American coast (Figure 1).

Methods

We trained a MaxEnt model on data from British Columbia using 19 environmental data layers, and extrapolated predictions for Alaska. We then trained 200 bootstrapped models, and calculated the standard deviation (SD) and the coefficient of variation (CV) of the predictions on a cell-by-cell basis for Alaska. We also applied the sum of sensitivity and specificity threshold to each bootstrapped model, and quantified the proportion of predictions falling in the least frequent class multiplied by two, and called this measure categorical uncertainty (CU; e.g., with 40 ‘presence’ and 160 ‘absence’ predictions, CU = 40/200 * 2 = 0.4). We then compared model predictions with the presence/absence data from Alaska, to show the relationship between estimated uncertainty (SD/CV/CU) and the actual accuracy of predictions.

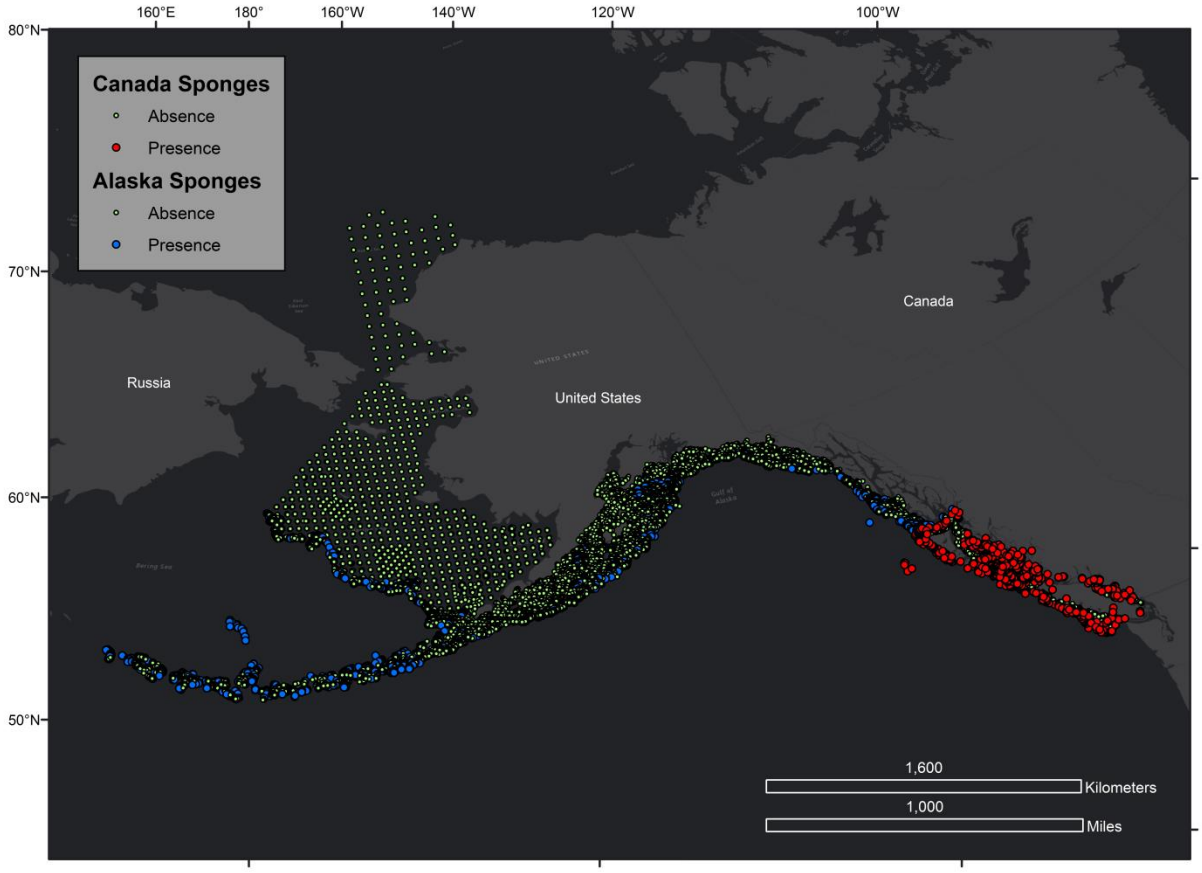


Figure 1. Sponge Data. Glass sponge presences in British Columbia, Canada (Red) and Alaska, USA (Blue), and absences (Green).

Results

Figure 2 shows the predictions made for the Alaska region, extrapolated using the MaxEnt model trained with data from British Columbia, as well as the three uncertainty metrics. Figure 3 shows the overall accuracy of the predictions, binned into 10 equally spaced ranges of uncertainty estimates.

Discussion

For the SD and CU metrics, data bins with higher uncertainty (on the right side of each plot) have lower accuracies, and vice versa, suggesting that these two metrics can effectively be used to quantify the amount of confidence that should be placed in the predictions. For the CV metric, on the other hand, data bins with high uncertainty have high accuracy of predictions, suggesting that this metric should not be used to guide confidence in predictions.

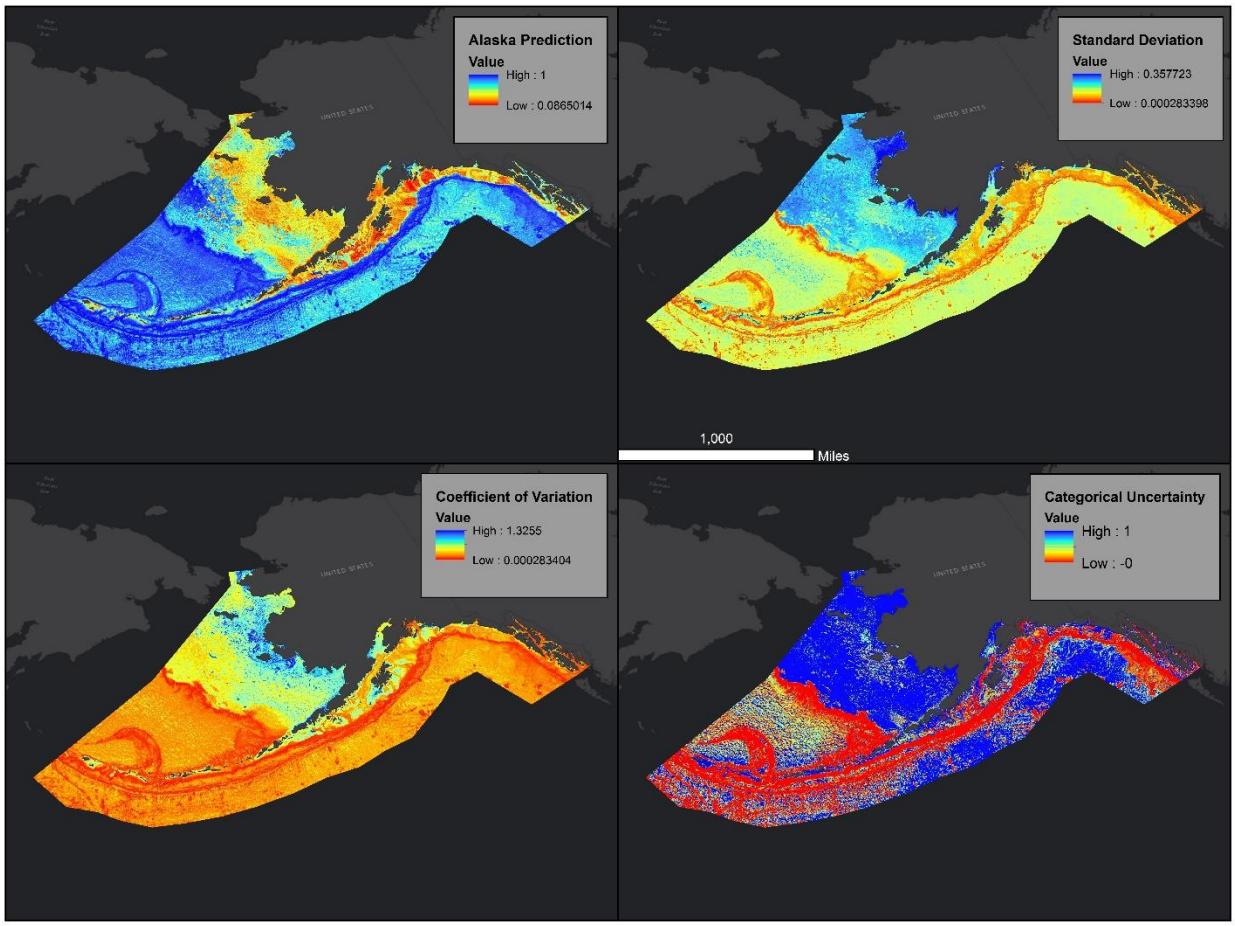


Figure 2. Top left: Mapped predictions of sponge presence/absence in Alaska. Top right: Standard deviation (SD) of bootstrapped predictions. Bottom left: Coefficient of variation (CV) of bootstrapped predictions. Bottom right: Categorical uncertainty (CU) based on bootstrapped predictions.

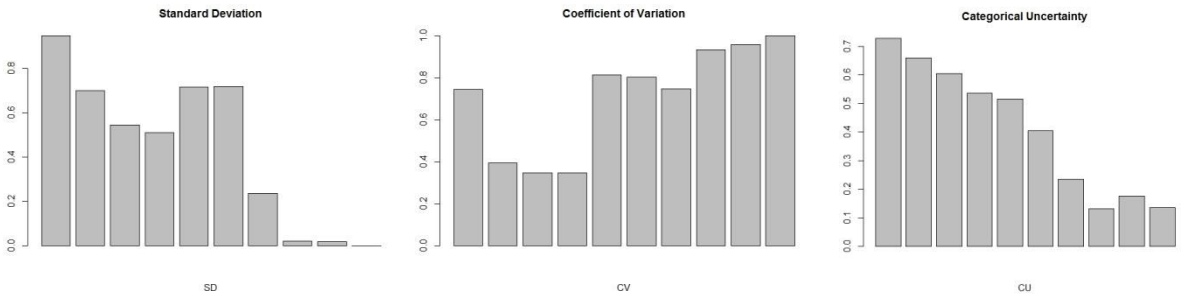


Figure 3. SD, CV and CU values, binned, and the corresponding accuracy of predictions.

References

Anderson, O.F., Guinotte, J.M., Rowden, A.A., Tracey, D.M., Mackay, K.A., and Clark, M.R. 2016. Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. Deep-Sea Res. Part I: Oceanogr. Res. Pap. 115: 265–292.

Rowden, A.A., Anderson, O.F., Georgian, S.E., Bowden, D.A., Clark, M.R., Pallentin, A., and Miller, A. 2017. High-Resolution Habitat Suitability Models for the Conservation and Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean. Front. Mar. Sci. 4: 335.

4. THEME 2 - BIOLOGICAL AND ENVIRONMENTAL DATASETS RELEVANT TO DEEP-SEA SPECIES AND COMMUNITIES

Chair: Dr. Telmo Morato

Issues related to data gaps and data quality hampering the development of improved species distribution models in the deep sea were discussed. This discussion focused largely on data quality, accuracy, extent, resolution, ecological relevance, and temporal scales.

Some of the points made are summarised below:

1. Overall data availability is improving. Archiving the available environmental data and derived data so that it is open access would be important moving forward.

2. Scientists should be encouraged to publish new environmental data in 'data journals' rather than as supplementary data linked to the original research paper to facilitate more ready access.

3. There is a need for a shared data repository, for example linked to an expert forum similar to the group assembled for this workshop. Data layers from many institutes could be continuously updated, however quality control and oversight would be required for such a project. Also, a critique of new papers coming out would also be helpful. A web portal could be created with metadata of relevant datasets. Users could rank their usefulness using a simple 1 to 5 star system. An example of an existing repository is BioOracle (http://www.bio-oracle.org). Action: Dr. Andrew Davies agreed to set up a simple website where relevant datasets could be listed in order to facilitate species distribution modeller's efforts.

4. Required data quality relates directly to the scale at which the model will be used.

Biological data needs to be georeferenced with much greater care and attention; often, very high-quality bathymetry is available for an area and the georeferencing of biological data becomes the weak point. Actual position reference data is required for samples, rather than data aggregated at a 1 km cell size. Often, global databases contain limited metadata describing what and how data were collected. Often only the geographic coordinates describing the location of the organism are given with no information about gear type/length of tow used to collect the sample available - this increases uncertainty. OBIS (Ocean Biogeographic Information System) facilitates entry of comprehensive metadata associated with biological data if available.

5. In terms of ecologically relevant data, for species such as corals and sponges, process studies are very important in identifying variables which are important to individual species. It was noted that adding hydrodynamic data into species distribution models will be an important next step (e.g., Mohn et al., 2014).

A short discussion on the topic of new data sources proposed the inclusion of detection probability. To do this, repeat sampling surveys would be needed, but this will only be possible for some species. Occupancy counts are often lower than actual occupancy due to the probability of missing some counts; this should be quantified in some way, e.g., using still images taken from the same sampling sites. For cryptic deep-water sponge species - repeat images could yield a value for detection probability. How this differs across the study area is also potentially significant.

The importance of biological traits was then discussed and how they might be incorporated into modeling approaches. A major issue identified was the availability of sufficient data. It was mentioned that information pertaining to other closely related species could be used to help fill gaps in knowledge for certain species (see e.g., the Marine Ecosystem Research Program (MERP); the Trait Explorer).

The Hierarchical Modeling of Species Communities (HMSC) package opens the door for including traits and modeling the response to covariates in the model.

1. Joint Species Distribution Models (JSDMs) are still rather new; it has been demonstrated that when trait information is included, the result is similar to responses of species to their environment. In some cases, it has been demonstrated as a positive influence on the model, but other cases show including species-to-species interactions also helps; this does not need to come from trait information. The deep sea is not the best place to test species interactions and their impact upon models, however linking ecosystem models with species models can yield interesting results.

2. Looking at connectivity and MPA design, SDMs can be used to understand population

structure and how it matches environmental gradients (looking at adaptations of specific groups). Genetic differences can be used as input to models rather than environmental data, and this can produce interesting habitat delineated areas. Suitability models can be developed for either side of the genetic break (the WoRMs repository can be used as a starting point). Species responses to climate change will vary based on their habitat determined genetic structure, so this is an important theme to follow up on.

Finally, the concept of community modeling was discussed (particularly the question of whether there is specific data required for this that has not been addressed).

1. Many deep-sea species have very wide dispersions. When regional studies are created, these don't consider their entire distribution; it is known that species exists outside of the area for which a local model is produced.

2. If there is a species present in different ocean basins, how should that data be combined so that location specific traits are not erased? Fisheries stock assessment models use joint stock/population distribution models. Each stock has its own parameters based on where it is located. Each regional population should be treated as having its own niche, and each population should be modeled rather than the whole species.

References

Mohn, C., Rengstorf, A., White, M., Duineveld, G., Fienis, F., Soetaert, K. and Grehan, A. 2014. Linking benthic hydrodynamics and cold-water coral occurrences: A high-resolution model study at three cold-water coral provinces in the NE Atlantic. Progress in Oceanography 122:92–104.

5. THEME 3 - TEMPORAL AND SPATIAL SCALES RELEVANT FOR DEVELOPING SDM IN THE DEEP SEA

Chair: Dr. Skipton Woolley

This discussion focused on issues of temporal scale as well as spatial scale. In terms of phylogenetic information, the discussion considered what kind of factors would be useful for species distribution models in a deep-sea context where the species may not be fully identified. A number of points were made:

1. It is possible to look at niche patterns across phylogeny with the assumption that closer species have similar niches. This information can be used to infer data for species for which little is known.

2. With regard to niche partitioning, species are very good at finding niches that have not been occupied previously, for example they inhabit different bathymetry to avoid competition or unfavourable environments. How can this knowledge be applied in models? The hypothesis that closer species have similar niches can be tested through species distribution models.

3. Exploring how species change with geography is an approach which can help interpret models. In one example of species archetype modeling, two resultant groups corresponded with phylogenetic history (Arctic vs Boreal sponges).

Short versus long-term climate variability. It is important to note that when using climate predictions (short-term) or climate projections (long-term), research questions might be fundamentally different between these two time scales. As well, the data types used will also be different.

1. Historical data can be used for validation of modelled temperature values. This can provide increased confidence in the prediction.

2. If the habitat distribution between a predator and prey (for example) changes with predictions, can this be modeled using JSDM? Yes, and scenario predictions can be used as well, for example, what would happen if one species was removed? It is more difficult to add predictions to JSDM and interpret these results.

3. What is a reasonable time frame to be looking at? For example, if data is available per five days for 20 years, or per month for 20 years. The purpose of the study will drive the answer to this question, for example, if the question is looking at shifting geographically it would be important to have a long time-frame.

Potential for adaptation to climate variability. Different data covariates can result in different model results. In order to make mechanistic models productive, information on microclimate is required. For example, when looking at how koalas survived in different heat scenarios, by looking at the microclimate it was possible to see how the koala would find the coolest area possible. This presents the issue of scale, is it possible to have fine-scale data such as this for the deep sea? Perhaps it would be more relevant to take for example *in situ* records of specifics such as water temperature and salinity at the exact location of species occurrence and use a spatial method of analysis. Values such as these can then inform spatial model creation.

Finally, the discussion focused on the main issues of spatial scales.

1. Probability of occurrence is less meaningful without information about the sampling unit. Recommendation for future work involves always being explicit

in what the probability of occurrence actually references and how it was calculated.

2. The occurrence probability is an average for the grid cell, physical features may be smoothed in the prediction maps, this means that the probability of occurrence would be very low because the majority of that cell would be unlikely habitat for that species. This needs to be made clear throughout the project.

3. When using data such as commercial fishing trawl occurrences, there is bias because these trawls will occur in highly productive areas as well as uncertainty in the data as trawl gear aggregates catches over long distances.

6. THEME 4 - MODELING TOOLS IN THE CONTEXT OF DATA-LIMITED SITUATIONS AND IN THE CONTEXT OF SINGLE SPECIES AND JOINT SPECIES MODELING

Chair: Dr. Andrew Davies

This discussion had three topics: issues with modeling in data-limited situations, future directions of modeling tools in these situations, and finally any suggestions for overall modeling approaches.

To begin, the main issues relating to modeling in data-limited situations were looked at. Some starting terms were: model uncertainty, model validation, modeling assumptions, model ensembling, model transferability, thresholds for interpretation of probabilistic maps.

1. In the instance of producing species distribution modeling work for guiding policy, there should be more weight placed on measuring uncertainty in the model output in order to not produce an overestimate.

2. Issues associated with calibration and validation were next presented:

    a. It is important to separate statistical validation and ecological validation;

    b. In terms of metrics, is AUC useful? Selecting covariates for the model is one of the most important steps, what form do they take (linear, unimodal etc.).

3. While an ensemble prediction/multi-model approach is often considered to be better, it must also be considered that the likelihood of finding one model that fits well with a specific case is higher. However, if doing ensemble prediction, the models must be weighted properly. One problem raised with ensemble prediction is that response curves are lost. Is the relationship between species not one of the most important aspects of modeling? The second problem with the multi-model approach is if the models have to

be weighted because some are bad, why include bad models in the ensemble prediction? A well-constructed single species model may give a better result than many less well-constructed models.

The second main topic discussed in this theme focused on if we can identify or suggest future direction or guidelines for modeling tools in the context of data-limited situations.

1. A unique approach would be to use something like BRT or MaxEnt (especially if presence only data is available) and use this to identify functional relationships. By looking at the output plots of these models, useful information is produced concerning the relationships. The problem with model types like BRT or MaxEnt is then the uncertainty measures are not the best. A multi-model approach would be best in this case, with the incorporation of GLMs or GAMs.

2. It is important to consider what ideally will be achieved and what data sources are available. Often it is impossible to rely on available data to be representative of the system, this is why a model-based approach should always be taken.

3. Using informative priors from literature as a basepoint to start models was a method suggested of moving forward with modeling in data-limited situations.

    a. By building a habitat suitability index before beginning the modeling, one can produce base level data that the model can then be compared to.

4. Finally, it was noted that machine learning depends on a lot of data, this is not available for the deep sea. If there is enough data available, it is significantly easier to reach a correct, or close-to-correct answer.

The final topic for this theme was focused on looking at future directions for modeling approaches in the deep sea:

1. It was noted that this depends on the audience as recommendations will vary between experts and non-experts.

2. Ideally there would be an increase in publications outlining new outlines for SDM work.

3. How should model performance be assessed, many current evaluation methods are perhaps not the most important. It may be best to continue to use GLM and GAM formats.

7. DEVELOPMENT OF GUIDELINES

Workshop participants divided into one of three break-out groups to formulate ideas on what type of content might go into the future development of *Guidelines for Species Distribution Modeling in the Deep Sea*. The notes from those discussions are summarized here in the format presented. Ideally these notes will be translated into a structured set of guidelines that can be published separately.

Biological data

1. What sort of data do you have?

    a. Good and well-understood dataset: systematic repeated survey; well randomised design; conducted systematically; good spatial coverage of area of interest; known and interpretable gear; quantifiable detectability; standard operation procedure.

    b. Ad hoc/OBIS/literature search: QC of these sorts of data is important, so need to consider - taxonomic alignment; can we assess spatial accuracy (spatial uncertainty); depth-based validation?; how was the data collected?; what are the data gaps?; What are the spatial/taxonomic/environmental/temporal biases?

2. Think about the model

    a. What are the objectives of the model? – used to examine global or local patterns, answer ecological questions or used as managements tools?

    b. Are the data good enough to model? - take pre-emptive steps to address before modeling; does the resolution of the response and environmental data match (particular problem for deep sea)?

Environmental data

1. Where are you looking at? - e.g., predicting suitable habitat distribution or connectivity and dispersal of organisms.

2. Think about the justification of your variable choice and how you can explain it – e.g., learn about the physiology of your modelled species - what environmental variables may be known that limit the distribution of your species?

3. Understand your environmental data before model use – e.g., consider scales of data; bathymetry - look at acoustic and satellite - where are interpolation?; what do derivatives of bathymetry mean for us?; check interpolation of your data.

4. Validate the environmental data if you can – e.g., do we have direct measurements - use it to validate - what is a realistic expectation of comparison; look at environmental profiles of your samples - is this what you expect for your species?

5. Special considerations for the deep sea? – e.g., bathymetry; bathymetry derivatives; surface layers - are they of relevance?; 3d models; flow direction importance?; proxies?; spatial bias; distance to seamount; internal waves?; human impacts?

6. Use layers of the same resolution where possible and consider whether up-scaling or down-scaling is appropriate

Modeling methods

1. What desirable qualities in a model do we want? – e.g., uncertainty; clarity of what the model is doing; does the model formally describe the data?; ease to use; reproducible (using scripts and makes the modeling steps very transparent); purpose of prediction; interpolation and extrapolation and report uncertainty; suitable for hypothesis testing.

2. Predicting outside environment space and geographical space - if we are willing to make a hypotheses about the prediction and explicitly state that hypotheses in modeling steps and model reporting.

For example, we may be willing to predict to the western Pacific Ocean from a model that is fit to eastern Pacific Ocean. Under the assumptions that the models fitted in the west are representative for the animals we find in the east. Providing these assumptions and hypotheses, we can justify model prediction and extrapolation into other regions. Another example might be the lessons that come from invasive species modeling and understanding the potential extrapolations of where species might be to understand risk for management objectives.

3. Something to do within a paper that describes guidelines – would be to use data degradation to demonstrate that for good data cases that there are readily made tools which do the model you want, but as the data degrades or becomes more complex, then you have to use more complicated modeling packages or other variations on these models to deal with the data. Then if it gets really complicated, you might have to create your own model and/or talk to a statistician and get some help with building a more complex model which can deal with nuances in the data you are modeling.

4. Do we need a spatial prediction? Based on the objectives of the outputs can we get more informative models based on a-spatial statistical models?

5. Identifying candidate models (your chance to talk to a statistician). Has the approach to model identification been justified with respect to published advice, statistician's advice, and the availability of species survey data? Is uncertainty formally reported within the model you are

choosing? Given the choice of the model what package should we use? e.g., Package 'ppmlasso' instead of MaxEnt.

6. When given a choice use a statistical-model based approach. For example, MaxEnt and IPPMs are the same thing, so use a IPPM which formally defines the model and removes the black box. Desirable qualities of statistical models – they describe the error distribution and explain it with respect to described model. Not a machine learning, but algorithmic approaches. There are many machine learning approaches which are based on statistical modeling.

7. Dependent and independent variables in model fitting and testing. Have the statistical relationships between environmental variables and species occupancy (or abundance) been provided in a clear graphical form? Visualising the response of coefficients during model fitting and test. Because of the sparse nature of data, we should try and fit ecological relevant relationships.

8. Model diagnostics. Have model diagnostics been presented and discussed (e.g., residuals)?

9. Statistical evaluation, model selection and cross-validation.

a. Has model evaluation been undertaken? Have output habitat maps produced by the 'best model' and the 'most competitive of the non-preferred plausible models been presented and considered? Has the relative statistical support for each model been presented using information criteria?

b. Cross-validation is the number one method for validating the fit of the model. When testing the model through cross validation use left out, and make sure your predictions from a training model are correctly fitting a testing model. Do model validation well and make sure you are happy with the predictions based on the data used to fit the model. Touched on the different cross-validation methods and what they can do and what they are useful for.

c. Model selection is very hard and no one accuracy metric is the best. Even the current evaluation methods such as AUC and TSS are controversial. When considering an accuracy metric its best to report several (e.g. AUC, TSS, Kappa) and discuss the pros and cons in relation to the type of error you want to control (e.g. omission or commission error).

.

d. Model comparison of different regions or dataset is ok, if you are honest with the potential issues with differences in the way the data was collected and species you are making inference about. You are the best expert about the data, you need to be critical if datasets are comparable. Leave out model-averaging because it's very hard to do well.

Guidelines for SDM products

1. Maps

a. People want maps in a form they can use. Should discuss with end-users, provide GIS immediately (colour ramps etc).

b. Maps showing uncertainty are necessary: coefficient of variance; confidence intervals (how close to estimating mean) or prediction intervals; include input data points (especially absence or pseudo-absence); areas of extrapolation – physical extent of extrapolation; show depth distribution of where samples were taken. Add histogram showing depth distribution of data points.

2. Thresholds

a. Both predicted probability (percentage) maps and binary maps showing areas of suitable vs. unsuitable habitat are useful depending on the application (e.g. conservation management).

b. Choose a percentage of the probability distribution function (e.g., 80%) and plot that rather than a threshold based on the overall probability.

c. Try to obtain ecological (in field) validation data for your thresholds (understand assumptions).

d. Can we establish minimum guidelines for field validation?

3. Data accessibility/archiving (for stakeholders)

a. Be prepared to make all underlying raw data and modelled data available (for transparency) to (project) stakeholders. Identify good repositories (e.g., Australian example (video and photo)).

b. Ensure long-term archiving of data. Local, regional and global data repositories (institutes should have automated process to archive data). More problematic for individual academic researchers. Australia – Integrated Marine Observing System has a portal called the Australia Ocean Data portal that takes all data. New Zealand example – will archive data in old publications etc.

4. Climate change predictions

a. Which are the most appropriate sources of climate change scenario information?

b. It's important to distinguish whether to use climate prediction (short time frame tuned to current climate conditions) versus climate projection (longer term projection – not tuned to current climate conditions). Need a rationale to choose the time frame used in modeling. Management decisions are typically required on a 10-20 year time frame.

c. Ecologists need a better understanding of how climate change models work.

d. Link predictions to follow-up monitoring to see which trajectory your ecosystem is on in climate change terms.

e. What parameters should be monitored? Annual ENSO/tuna catch/tuna futures. Depends on species (mobile vs sedentary).

f. Climate change SDMs may indicate direction of migration of species (e.g., towards poles, deeper or shallower) or identify areas of increased conservation importance in the future due to contractions/concentration of species. Implications for MPA design. May need new MPA areas to buffer future species movements.

g. Validate niches identified in climate change prediction models with measured data on environmental tolerances when possible.

5. Modeling Decision Support Tree

a. Work being undertaken by Jessica Finney for DFO and should be available in early 2019.

8. ACKNOWLEDGEMENTS

Participation of the keynote speakers was funded by Fisheries and Oceans, Canada, Maritimes Region. This workshop has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nos. 678760 (ATLAS) and 679849 (SponGES). This output reflects only the author's views and the European Union cannot be held responsible for any use that may be made of the information contained therein. We additionally thank the Office of the Secretariat of the Convention on Biological Diversity (CBD), and in particular Jihyun Lee, Joseph Appiott, Changsung Lim and Johany Martinez for allowing the meeting to take place in their meeting room, assisting us with logistics and for their interest and support for our work.

APPENDIX 1. WORKSHOP AGENDA

WORKSHOP ON THE USE OF SPECIES DISTRIBUTION MODELING IN THE DEEP SEA

May 17-18, 2018 - MONTREAL, QUEBEC, CANADA

Venue: Joke Waller-Hunter Conference Room, Secretariat of the Convention on Biological Diversity, 413 Saint-Jacques Street, Suite 800, Montreal, Quebec H2Y 1N9

Thursday, 17 May

815	Meet in the Lobby for security escort to meeting room
840	Convenors Welcome **(Ellen Kenchington)**
845	Welcome from the CBD
900	**Setting the Context: Voluntary Specific Workplan on Biodiversity in Cold-water Areas within the Jurisdictional Scope of the Convention** Murray Roberts, ATLAS Project Co-ordinator, University of Edinburgh
915	**Theme 1 Showcase Approaches to Develop SDM/HSM in the Deep-sea (Anthony Grehan)**
920	**Keynote Talk: Dr Piers Dunstan, CSIRO Marine Biodiversity Hub, Australia**
1030	Coffee/Tea Break and Mixer
1100	**Keynote Talk: Dr Jarno Vanhatalo, University of Helsinki, Finland**
1145	Open Discussion of Keynote Talks
1230	Lunch (Catered In) to Facilitate Informal Discussions
1330-1730	**Participant's Forum (Ellen Kenchington Moderator)** **No formal Coffee Break but coffee will be available through the afternoon (self service)** Short presentations by participants on pertinent issues collated by subject followed by discussion (30 min approximately) 1. What do environmental managers and stakeholders want from a model? (Ashley Rowden) 2. Resolution of seabed features in the deep-sea: Implications for habitat characterization. (Myriam Lacharite) 3. Determining thresholds for interpretation of probabilistic maps. (Chris Rooper) 4. Model validation. (Chris Rooper) 5. Evaluating effects of rescaling and weighting data on habitat suitability modeling. (Kisei Tanaka) 6. How to undertake ensembling of multiple models/multiple data sets. (Chris Rooper) 7. Distribution models as a tool to predict climate change effects in the deep sea. (José González)

8. A generalized climate-niche modeling framework for commercially important species: A case study with American lobster and Atlantic scallop in the Gulf of Maine. (Kisei Tanaka)

1900 Group Dinner Restaurant Les Pyrénées, 320 Saint-Paul St W, Montreal

Friday, 18 May

845 **Theme 2 Biological and Environmental Datasets relevant to Deep-sea Species and Communities (Telmo Morato)**

1030 Coffee/Tea Break with Mixer

1100 **Theme 3 Temporal and Spatial Scales Relevant for Developing SDM/HSM in the Deep-Sea (Skip Woolley)**

1230 Lunch (Catered In) to Facilitate Informal Discussions

1330 **Theme 4 Modeling Tools in the Context of Data-Limited Situations and in the Context of Single Species and Joint Species Modeling (Andy Davies)**

1430 **Development of Guidelines (Chris Yesson and Anthony Grehan)**
Break out groups

1630-1700 **Conclusions and Next Steps (Ellen Kenchington)**

Notes for Participants

Participants have been asked to prepare short (max. 1-2 pages) working papers addressing agenda items and to make short presentations (10 min max – 2-4 ppt slides) to stimulate discussion. These do not necessarily have to be on the same topic. The workshop will address data, knowledge, methods, and performance gaps and identify further steps to improve species distribution modeling in the deep sea. However, we also want to use this opportunity for participants to discuss problems that they have encountered or areas where they would appreciate other viewpoints. Therefore, the working paper could provide background for an issue that will be discussed in the presentation, or it could be a stand-alone paper on recent work. The presentations will slot into the Participants Forum or one of the Theme Sessions depending upon the topic. There is no requirement to make a presentation but we hope that you will!

Timelines: (please send to Ellen.Kenchington@dfo-mpo.gc.ca)

Presentation Topic	Before 1 April 2018
Working Papers	30 April 2018
PPT Presentation	Before 15 May 2018

Publications

The working papers will be combined along with the meeting notes and conclusions into a jointly authored technical report in the Canadian Technical Report of Fisheries and Aquatic Sciences series (http://publications.gc.ca/site/eng/9.514691/publication.html).

Participants may also want to consider publishing their work in the ‘Managing Deep-sea Ecosystems at Ocean Basin Scale’ Research Topic that Telmo Morato and Murray Roberts are putting together for Frontiers in Marine Science: https://www.frontiersin.org/research-topics/7768/managing-deep-sea-ecosystems-at-ocean-basin-scale . The deadline for abstracts is 12 July 2018 with the manuscript due 15 November 2018. This is something we can discuss at the workshop and if some of the topics warrant it we could look at putting together collaborative (or otherwise) papers to submit.

Dropbox

A dropbox will be set up to allow participants to upload papers (their own or others) that bear on the workshop. Please name files using the following format: FIRST AUTHOR SURNAME_YEAR_JOURNAL ABBREV and check to see that the paper has not already been posted.

Travel Notice

Non-Canadian travellers may need an Electronic Travel Authorization (eTA) to board their flight to Canada. It is the responsibility for each traveller to obtain this (it is not part of the airline ticket). For more information on how to purchase this please visit the links provided in the Information package of 9 March 2018. *The cost is very small and can be done online.*

APPENDIX 2. LIST OF PARTICIPANTS

Participant	Affiliation	Email Address
Appiott, Joseph	CBD, Montreal, PQ Canada	joseph.appiott@cbd.int
Beazley, Lindsay	DFO, BIO, Canada	Lindsay.Beazley@dfo-mpo.gc.ca
Callery, Oisin	NUI, Galway, Ireland	oisin.callery@nuigalway.ie
Davidson, Fiona	University of Ottawa, Ontario, Canada	fdavi065@uottawa.ca
Davis, Andrew	University of Bangor, Bangor, UK	andrew.j.davies@bangor.ac.uk
Dunstan, Piers	CSIRO, Oceans and Atmosphere, Hobart, Tasmania, Australia	Piers.Dunstan@csiro.au
Du Preez, Cherisse	DFO, Pacific Region, Canada	Cherisse.DuPreez@dfo-mpo.gc.ca
Finney, Jessica	DFO, Pacific Region, Canada	jessica.finney@dfo-mpo.gc.ca
González-Irusta, José	IMAR, Azores, Portugal	gonzalezirusta@gmail.com
Grehan, Anthony	NUI, Galway, Ireland	Anthony.Grehan@nuigalway.ie
Howell, Kerry*	University of Plymouth, Plymouth, UK	kerry.howell@plymouth.ac.uk
Kenchington, Ellen	DFO, BIO, Canada	Ellen.Kenchington@dfo-mpo.gc.ca
Knudby, Anders	University of Ottawa, Ontario, Canada	aknudby@uottawa.ca
Lacharité, Myriam	Nova Scotia Community College, Dartmouth, NS, Canada	myriam.lacharite@nscc.ca
Lee, Jihyun	CBD, Monteal, PQ Canada	jihyun.lee@cbd.int
Morato, Telmo	IMAR, Azores, Portugal	t.morato@gmail.com
Murillo, Javier	DFO, BIO, Canada	Javier.MurilloPerez@dfo-mpo.gc.ca
Roberts, Martyn	University of Bangor, Bangor, UK	emyr.roberts@bangor.ac.uk
Roberts, J. Murray	University of Edinburgh, Edinburgh, UK	Murray.Roberts@ed.ac.uk
Rooper, Chris	Alaska Fisheries Science Center, NMFS, NOAA, Seattle, Washington USA	Chris.Rooper@dfo-mpo.gc.ca
Rowden, Ashley	National Institute of Water and Atmospheric Research, Wellington, NZ	Ashley.Rowden@niwa.co.nz
Rubidge, Emily	DFO, Pacific Region, Canada	Emily.Rubidge@dfo-mpo.gc.ca
Stanley, Ryan*	DFO, BIO, Canada	Ryan.Stanley@dfo-mpo.gc.ca
Stirling, David*	Marine Scotland Science, Aberdeen, UK	D.A.Stirling@MARLAB.AC.UK
Tanaka, Kisei	Atmospheric and Oceanic Sciences Program, Princeton University, USA	kiseit@princeton.edu
Vanhatalo, Jarno	University of Helsinki, Helsinki, Finland	jarno.vanhatalo@helsinki.fi
Weigel, Benjamin	University of Helsinki, Helsinki, Finland	benjamin.weigel@helsinki.fi
Woolley, Skipton	CSIRO, Oceans and Atmosphere, Hobart, Tasmania, Australia	Skip.Woolley@csiro.au
Yesson, Chris	Zoological Society of London, London, UK	chris.yesson@ioz.ac.uk

*Indicates Webex Participation